Nonparametric Preference Completion: Supplementary Material

Julian Katz-Samuels and Clayton Scott
{jkatzsam,clayscot}@umich.edu

Department of Electrical Engineering and Computer Science, University of Michigan

A Outline

In Section B, we give the counterexample establishing Proposition 1 and give theorem proofs for the continuous rating setting. In Section C, we give theorem proofs for the discrete rating setting. In Section D, we prove the lemmas used in our theorem proofs, beginning with lemmas common to both the continuous rating setting and discrete rating setting and, then, presenting the lemmas on the continuous rating setting and discrete rating setting, separately. In Section E, we provide the proofs of the necessary and sufficient conditions. In Section F, we prove Proposition 2 and that the models $f(x, y) = x^t y$ and $f(x, y) = \|x - y\|_2$ are equivalent by adding a dimension. Finally, in Section G, we give some bounds that we use in the proofs for reference.

Unless otherwise indicated, all probability statements are with respect to $\{\boldsymbol{x}_i\}_{i \in [n_1]} \cup \{\boldsymbol{y}_u\}_{u \in [n_2]} \cup \boldsymbol{\Omega}$ in the continuous ratings setting and with respect to $\{\boldsymbol{x}_i\}_{i \in [n_1]} \cup \{\boldsymbol{y}_u\}_{u \in [n_2]} \cup \{\boldsymbol{a}_{u,l}\}_{u \in [n_2], l \in [L-1]} \cup \boldsymbol{\Omega}$ in the discrete ratings setting.

B Proofs for Section 5.1

To begin, we introduce some additional notation. When $\boldsymbol{y}_u$ and $\boldsymbol{y}_v$ are random, we write $R_{\boldsymbol{u},\boldsymbol{v}}$ instead of $R_{u,v}$ for emphasis.

Proof of Proposition 1. Consider the functions

$$f(z) = \begin{cases} \epsilon z & : z \in [0, \frac{1}{2}] \\ \epsilon(1 - z) & : z \in (\frac{1}{2}, 1] \end{cases}$$

and

$$g(z) = \begin{cases} -\epsilon z & : z \in [0, \frac{1}{2}] \\ \epsilon(z - 1) & : z \in (\frac{1}{2}, 1] \end{cases}$$

□

Next, we analyze Pairwise-Rank (PR), bounding the probability that Pairwise-Rank cannot distinguish between items i and j when $|f(x_i, y_u) - f(x_j, y_u)| > \epsilon$, i.e., the event

$$\begin{aligned} D^{\epsilon}_{u,i,j} \coloneqq \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) + \epsilon < f(\boldsymbol{x}_j, \boldsymbol{y}_u)\} \cap \{\mathrm{PR}(u, i, j, \beta, k) = 1)\}) \\ \cup \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) > f(\boldsymbol{x}_j, \boldsymbol{y}_u) + \epsilon\} \cap \{\mathrm{PR}(u, i, j, \beta, k) = 0\}). \end{aligned}$$

Theorem B.1. *Suppose* $\forall u \in [n_2]$, $g_u(z)$ *is strictly increasing. Let* $\epsilon, \delta > 0$ *and* $\eta \in (0, \frac{\epsilon}{2})$. *Suppose that almost every* $y \in \mathcal{Y}$ *is* $(\frac{\epsilon}{2}, \delta)$*-discriminative. Let* r *be a positive nondecreasing function such that* $r(\frac{\epsilon}{2}) \geqslant \delta$ *and* $r(\eta) < \frac{\delta}{2}$. *Suppose that almost every* $y \in \mathcal{Y}$ *is* r*-discerning. Let* $0 < \alpha < \frac{1}{2}$. *If* $p \geqslant \max(n_1^{-\frac{1}{2}+\alpha}, n_2^{-\frac{1}{2}+\alpha})$, $n_1 p^2 \geqslant 16$, *and* n_2 *is sufficiently large, for all* $u \in [n_2]$ *and* $i \neq j \in [n_1]$, *the output of Pairwise-Rank with* $k = 1$ *and* $\beta = \frac{p^2 n_1}{2}$ *is such that*

$$\begin{aligned} \Pr_{\{\boldsymbol{x}_i\},\{\boldsymbol{y}_u\},\boldsymbol{\Omega}}(D^{\epsilon}_{u,i,j}) \leqslant 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1 p^2}{8}) \\ + \exp(-(\frac{(n_2-1)p^2}{2})\tau(\eta)) + 3(n_2-1)p^2\exp(-\frac{\delta^2 n_1 p^2}{20}). \end{aligned}$$

The structure of the proof of Theorem B.1 is similar to the proof of Theorem 1 from Lee et al. (2016). The lemmas are distinct, however.

Proof of Theorem B.1. Fix $u \in [n_2]$, $i, j \in [n_1]$ such that $i \neq j$. Define:

$$W^{i,j}_u(\beta) = \{v \in [n_2] : |N(u,v)| \geqslant \beta, (i,v), (j,v) \in \Omega\}.$$

Further, define the events:

$$\begin{aligned} A &= \{|W^{i,j}_u(\beta)| \in [\frac{(n_2-1)p^2}{2}, \frac{3(n_2-1)p^2}{2}]\}, \\ B &= \{\max_{v \in W^{i,j}_u(\beta)} \rho(\boldsymbol{y}_u, \boldsymbol{y}_v) \geqslant 1 - \frac{\delta}{2}\}, \\ C &= \{|R_{\boldsymbol{uv}} - \rho(\boldsymbol{y}_u, \boldsymbol{y}_v)| \leqslant \frac{\delta}{4}, \forall v \in W^{i,j}_u(\beta)\}. \end{aligned}$$

By several applications of the law of total probability, we have that

$$\begin{aligned} \Pr(D^{\epsilon}_{u,i,j}) &= \Pr(D^{\epsilon}_{u,i,j}|A,B,C)\Pr(A,B,C) + \Pr(D^{\epsilon}_{u,i,j}|(A\cap B\cap C)^c)\Pr((A\cap B\cap C)^c) \\ &\leqslant \Pr(D^{\epsilon}_{u,i,j}|A,B,C) + \Pr(A^c) + \Pr((A\cap B\cap C)^c|A) \\ &\leqslant \Pr(D^{\epsilon}_{u,i,j}|A,B,C) + \Pr(A^c) + \Pr(B^c|A) + \Pr(C^c|A,B). \end{aligned}$$

We will upper bound each term in the above bound. By Lemma D.7, $\Pr(D^{\epsilon}_{u,i,j}|A,B,C) = 0$.

Setting $\lambda = \frac{1}{2}$ in Lemma D.1 yields that

$$\begin{aligned} \Pr(A^c) &= \Pr(|W^{i,j}_u(\beta)| \notin [\frac{(n_2-1)p^2}{2}, \frac{3(n_2-1)p^2}{2}]) \\ &\leqslant 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1 p^2}{8}). \end{aligned}$$

Lemma D.5 yields that

$$
\begin{aligned}
\Pr(B^c|A) = \Pr(\max_{v \in W_u^{i,j}(\beta)} \rho(\boldsymbol{y}_u, \boldsymbol{y}_v) < 1 - \frac{\delta}{2}|A) &\leqslant \Pr(\max_{v \in W_u^{i,j}(\beta)} \rho(\boldsymbol{y}_u, \boldsymbol{y}_v) < 1 - r(\eta)|A) \\
&\leqslant [1 - \tau(\eta)]^{\frac{(n_2-1)p^2}{2}} \qquad (1) \\
&\leqslant \exp(-(\frac{(n_2-1)p^2}{2})\tau(\eta)). \qquad (2)
\end{aligned}
$$

Line (1) follows by Lemma D.5 since conditional on A, $W_u^{i,j}(\beta) \geqslant \frac{(n_1-1)p^2}{2}$ and line (2) follows by the inequality $1 - x \leqslant \exp(-x)$. Since by hypothesis $\alpha \in (0, \frac{1}{2})$ is fixed such that $p \geqslant \max(n_1^{-\frac{1}{2}+\alpha}, n_2^{-\frac{1}{2}+\alpha})$, there exists a sufficiently large n_2 such that line (2) is less than $\frac{1}{2}$. Then, by Bayes rule, the union bound, and Lemma D.6,

$$
\begin{aligned}
\Pr(C^c|A,B) \leqslant \frac{\Pr(C^c|A)}{\Pr(B|A)} &\leqslant 2\Pr(C^c|A) \\
&= 2\Pr(\exists v \in W_u^{i,j}(\beta), |R_{\boldsymbol{uv}} - \rho(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\delta}{4}|A) \\
&\leqslant 3(n_2-1)p^2 \exp(-\frac{\delta^2}{4}\left\lfloor \frac{\beta}{2} \right\rfloor) \\
&= 3(n_2-1)p^2 \exp(-\frac{\delta^2}{4}\left\lfloor \frac{n_1 p^2}{4} \right\rfloor) \\
&\leqslant 3(n_2-1)p^2 \exp(-\frac{\delta^2 n_1 p^2}{20})
\end{aligned}
$$

where the last line follows because $n_1 p^2 \geqslant 16$ and $\forall x \geqslant 16$, $\lfloor \frac{x}{4} \rfloor \geqslant \frac{x}{5}$. Putting it all together, we have

$$
\begin{aligned}
\Pr(D_{u,i,j}^{\epsilon}) \leqslant\ & 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1 p^2}{8}) \\
&+ \exp(-(\frac{(n_2-1)p^2}{2})\tau(\eta)) + 3(n_2-1)p^2\exp(-\frac{\delta^2 n_1 p^2}{20})
\end{aligned}
$$

□

Proof of Theorem 1. For any $u \in [n_2]$, $i \neq j \in [n_1]$, define the event

$$
\begin{aligned}
\mathrm{Error}_{u,i,j}^{\epsilon} = &(\{f(\boldsymbol{x}_i, \boldsymbol{y}_u) + \epsilon < f(\boldsymbol{x}_j, \boldsymbol{y}_u)\} \cap \{A_{u,i,j} = 1\}) \\
&\cup (\{f(\boldsymbol{x}_i, \boldsymbol{y}_u) > f(\boldsymbol{x}_j, \boldsymbol{y}_u) + \epsilon\} \cap \{A_{u,i,j} = 0\}).
\end{aligned}
$$

Suppose that there exists $u \in [n_2]$ and distinct $i, j \in [n_1]$ such that $\mathrm{Error}_{u,i,j}^{\epsilon}$ occurs. Without loss of generality suppose that $f(\boldsymbol{x}_i, \boldsymbol{y}_u) + \epsilon < f(\boldsymbol{x}_j, \boldsymbol{y}_u)$, and $A_{u,i,j} = 1$. Then, inspection of the Multi-Rank algorithm reveals that $1 = A_{u,i,j} = \text{Pairwise-Rank}(u, i, j, \beta, k)$. Thus, $D_{u,i,j}^{\epsilon}$ occurs.

Therefore, by Theorem C.1 and the union bound,

$$\begin{aligned}
&\Pr(\exists u \in [n_2], i \neq j \in [n_1] \text{ s.t. } \mathrm{Error}^{\epsilon}_{u,i,j}) \\
&\quad \leqslant \Pr(\exists u \in [n_2], i \neq j \in [n_1] \text{ s.t. } D^{\epsilon}_{u,i,j}) \\
&\quad \leqslant n_2 \binom{n_1}{2} [2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1 p^2}{8}) \\
&\quad + \exp(-(\frac{(n_2-1)p^2}{2})\tau(\eta)) + 3(n_2-1)p^2 \exp(-\frac{\delta^2 n_1 p^2}{20})].
\end{aligned}$$

Now, suppose that $\forall u \in [n_2]$ and $i, j \in [n_1]$ such that $i \neq j$, $(\mathrm{Error}^{\epsilon}_{u,i,j})^c$ occurs. Then, by Lemma D.2, $\widehat{\sigma} = (\widehat{\sigma}_1, \ldots, \widehat{\sigma}_{n_2})$ with $\widehat{\sigma}_u = \mathrm{Copeland}(A_{u,:,:})$ satisfies $\mathrm{dis}_{2\epsilon}(\widehat{\sigma}, H) = 0$. □

Proof of Corollary 1. Ignoring constants, the two dominant terms in the bound in Theorem 1 are of the form $n_1^2 n_2 \exp(-n_2 p^2)$ and $n_1^2 n_2^2 \exp(-n_1 p^2)$. Then, under the conditions of Theorem B.1, as $n_2 \longleftarrow \infty$

$$\begin{aligned}
n_1^2 n_2 \exp(-n_2 p^2) &\leqslant \exp(2\log(n_1) + \log(n_2) - n_2^{2\alpha}) \\
&\leqslant \exp((1 + 2C_1)\log(n_2) - n_2^{2\alpha}) \longrightarrow 0.
\end{aligned}$$

Now, observe that

$$\begin{aligned}
n_1^2 n_2^2 \exp(-n_1 p^2) &= \exp(2\log(n_2) + 2\log(n_1) - n_1 p^2) \\
&\leqslant \exp(2\log(n_2) + 2\log(n_1) - n_1^{2\alpha}) \\
&\leqslant \exp(4\max(\log(n_2), \log(n_1)) - n_1^{2\alpha})
\end{aligned}$$

Suppose that $n_1 \geqslant n_2$. Then, clearly, the limit of the RHS as $n_2 \longrightarrow \infty$ is 0. Now, suppose that $n_1 < n_2$. Then, if $C_2^{2\alpha} > 4$, then as $n_2 \longrightarrow \infty$,

$$\begin{aligned}
n_2^2 n_1^2 \exp(-n_1 p^2) &\leqslant \exp(4\log(n_2) - n_1^{2\alpha}) \\
&\leqslant \exp([4 - C_2^{2\alpha}]\log(n_2)) \longrightarrow 0.
\end{aligned}$$

□

C Proofs for Section 5.2

To begin, because the model for the discrete ratings section is different, we introduce new notation in the interest of clarity. Fix $y_u, y_v \in \mathcal{Y}$. Define

$$\rho'(y_u, y_v) = \Pr_{\boldsymbol{g}_u, \boldsymbol{g}_v, \boldsymbol{x}_s, \boldsymbol{x}_t}[\boldsymbol{g}_u(f(\boldsymbol{x}_s, y_u)) - \boldsymbol{g}_u(f(\boldsymbol{x}_t, y_u))][\boldsymbol{g}_v(f(\boldsymbol{x}_s, y_v)) - \boldsymbol{g}_v(f(\boldsymbol{x}_t, y_v))] \geqslant 0).$$

Note that in this setting, the meaning of (ϵ, δ)-discriminative is slightly different.

Definition C.1. *Fix $y \in \mathcal{Y}$. Let $\epsilon, \delta > 0$. We say that y is* (ϵ, δ)-discriminative *if $z \in B_\epsilon(y)^c$ implies that $\rho'(y, z) < 1 - \delta$.*

In a sense, the notion requires in addition that the distribution of the monotonic functions reveals some differences in the preferences of the users.

Unless otherwise indicated, all probability statements are with respect to $\{\boldsymbol{x}_i\}_{i\in[n_1]} \cup \{\boldsymbol{y}_u\}_{u\in[n_2]} \cup \{\boldsymbol{a}_{u,l}\}_{u\in[n_2],l\in[L-1]} \cup \boldsymbol{\Omega}$. Next, we prove a theorem that is analogous to Theorem B.1. Recall the notation:

$$D^{\epsilon}_{u,i,j} \coloneqq (\{f(\boldsymbol{x}_i,\boldsymbol{y}_u)+\epsilon < f(\boldsymbol{x}_j,\boldsymbol{y}_u)\} \cap \{\mathrm{PR}(u,i,j,\beta,k)=1\}) \\ \cup (\{f(\boldsymbol{x}_i,\boldsymbol{y}_u) > f(\boldsymbol{x}_j,\boldsymbol{y}_u)+\epsilon\} \cap \{\mathrm{PR}(u,i,j,\beta,k)=0\}).$$

Theorem C.1. *Let $\epsilon,\delta>0$ and $\eta \in (0,\frac{\epsilon}{4})$. Suppose that $\mathcal{P}_{\mathcal{G}}$ is diverse and that almost every $y\in\mathcal{Y}$ is $(\frac{\epsilon}{4},\delta)$-discriminative. Let r be a positive nondecreasing function such that $r(\frac{\epsilon}{4})\geqslant\delta$ and $r(\eta)<\frac{\delta}{2}$. Suppose that almost every $y\in\mathcal{Y}$ is r-discerning. Let $\frac{1}{2}>\alpha>\alpha'>0$. If $p\geqslant\max(n_1^{-\frac{1}{2}+\alpha}, n_2^{-\frac{1}{2}+\alpha})$, $n_1p^2\geqslant 16$, $n_1\geqslant C_1\log(n_2)^{\frac{1}{2\alpha}}$ for some suitable universal constant C_1, and n_2 is sufficiently large, for all $u\in[n_2]$ and $i\neq j\in[n_1]$, the output of Pairwise-Rank with $k=n_2^{\alpha'}$ and $\beta=\frac{p^2n_1}{2}$ is such that*

$$\begin{aligned}\Pr_{\{\boldsymbol{x}_i\},\{\boldsymbol{y}_u\},\{\boldsymbol{a}_{u,l}\},\boldsymbol{\Omega}}(D^{\epsilon}_{u,i,j}) \leqslant & 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1p^2}{8}) + 2\exp(-\gamma(\frac{\epsilon}{4})k) \\ &+\frac{1}{1-r(\frac{\epsilon}{2})}[3(n_2-1)p^2\exp(-\frac{\delta^2n_1p^2}{20}) \\ &+\exp([1-\kappa(\frac{\epsilon}{2})+\tau(\eta)+\log(\frac{3(n_2-1)p^2}{2})]k \\ &-k\log(k)-\tau(\eta)\frac{(n_2-1)p^2}{2})].\end{aligned}$$

Proof of Theorem C.1. Fix $u\in[n_2]$, $i,j\in[n_1]$ such that $i\neq j$. Define:

$$W_u^{i,j}(\beta)=\{v\in[n_2]:|N(u,v)|\geqslant\beta,(i,v),(j,v)\in\Omega\}.$$

Further, define the events:

$$\begin{aligned}A&=\{|W_u^{i,j}(\beta)|\in[\frac{(n_2-1)p^2}{2},\frac{3(n_2-1)p^2}{2}]\},\\ B&=\{\max^{(k)}_{v\in W_u^{i,j}(\beta)}\rho'(\boldsymbol{y}_u,\boldsymbol{y}_v)\geqslant 1-\frac{\delta}{2}\},\\ C&=\{|R_{\boldsymbol{uv}}-\rho'(\boldsymbol{y}_u,\boldsymbol{y}_v)|\leqslant\frac{\delta}{4},\ \forall v\in W_u^{i,j}(\beta)\}\\ E&=\{|f(\boldsymbol{x}_i,\boldsymbol{y}_u)-f(\boldsymbol{x}_j,\boldsymbol{y}_u)|>\epsilon\}\\ M&=\{\exists v\in W_u^{i,j}(\beta)\text{ s.t. }\rho'(\boldsymbol{y}_u,\boldsymbol{y}_v)\geqslant 1-\frac{\delta}{2}\text{ and }\exists l\in[L-1]\text{ s.t. }\boldsymbol{a}_{v,l}\in(f(\boldsymbol{x}_j,\boldsymbol{y}_v),f(\boldsymbol{x}_i,\boldsymbol{y}_v))\}\end{aligned}$$

By several applications of the law of total probability, we have that

$$\begin{aligned}\Pr(D^{\epsilon}_{u,i,j}) &\leqslant \Pr(D^{\epsilon}_{u,i,j}|E) + \Pr(D^{\epsilon}_{u,i,j}|E^c)\\ &= \Pr(D^{\epsilon}_{u,i,j}|E)\\ &\leqslant \Pr(D^{\epsilon}_{u,i,j}|A,B,C,M,E) + \Pr(A^c|E) + \Pr(B^c|A,E)\\ &+ \Pr(C^c|A,B,E) + \Pr(M^c|A,B,C,E)\\ &= \Pr(D^{\epsilon}_{u,i,j}|A,B,C,M,E) + \Pr(A^c) + \Pr(B^c|A,E) \qquad (3)\\ &+ \Pr(C^c|A,B,E) + \Pr(M^c|A,B,C,E)\end{aligned}$$

Line (3) follows from the independence of $\boldsymbol{\Omega}$ from $\{\boldsymbol{x}_s\}_{s\in[n_1]}$ and $\{\boldsymbol{y}_v\}_{v\in[n_2]}$. We will bound each term in the above upper bound. By Lemma D.12,

$$\Pr(D^{\epsilon}_{u,i,j}|A,B,C,M,E) = 0. \qquad (4)$$

Setting $\lambda = \frac{1}{2}$ in Lemma D.1 yields that

$$\begin{aligned}\Pr(A^c) &= \Pr(|W^{i,j}_u(\beta)| \notin [\frac{(n_2-1)p^2}{2}, \frac{3(n_2-1)p^2}{2}])\\ &\leqslant 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1 p^2}{8}). \qquad (5)\end{aligned}$$

Next, we bound $\Pr(B^c|A,E)$. By Bayes theorem,

$$\begin{aligned}\Pr(B^c|A,E) &\leqslant \frac{\Pr(B^c|A)}{\Pr(E|A)}\\ &= \frac{\Pr(B^c|A)}{\Pr(E)} \qquad (6)\\ &< \frac{\Pr(B^c|A)}{1-r(\frac{\epsilon}{2})}. \qquad (7)\end{aligned}$$

Line (6) follows from the independence of $\boldsymbol{\Omega}$ from $\{\boldsymbol{x}_s\}_{s\in[n_1]}$ and $\{\boldsymbol{y}_v\}_{v\in[n_2]}$. Line (7) follows since by hypothesis almost every $y \in \mathcal{Y}$ is r-discerning.

Since almost every $y \in \mathcal{Y}$ is $(\frac{\epsilon}{4}, \delta)$-discriminative and r-discerning, and $\eta > 0$ is such that $r(\eta) < \frac{\delta}{2}$, Lemma D.9 yields that

$$\begin{aligned}\Pr(B^c|A) &= \Pr(\max^{(k)}_{v\in W^{i,j}_u(\beta)}\rho'(\boldsymbol{y}_u,\boldsymbol{y}_v) < 1-\frac{\delta}{2}|A)\\ &\leqslant \Pr(\max^{(k)}_{v\in W^{i,j}_u(\beta)}\rho'(\boldsymbol{y}_u,\boldsymbol{y}_v) < 1-r(\eta)|A)\\ &\leqslant \exp((1-\kappa(\frac{\epsilon}{4}) + \tau(\eta) + \log(3\frac{(n_2-1)p^2}{2}))k - k\log(k) - \tau(\eta)\frac{(n_2-1)p^2}{2})). \qquad (8)\end{aligned}$$

Next, we bound $\Pr(C^c|A,B,E)$. By Bayes theorem,

$$\Pr(C^c|A,B,E) \leqslant \frac{\Pr(C^c|A,B)}{\Pr(E|A,B)}.$$

Fix $\boldsymbol{y}_u = y_u$ r-discerning such that A and B occur. Then, since $\{\boldsymbol{x}_s\}_{s\in[n_1]}$, $\{\boldsymbol{y}_v\}_{v\in[n_2]}$, and $\boldsymbol{\Omega}$ are independent and y_u is r-discerning,

$$\begin{aligned}
&\Pr_{\{\boldsymbol{y}_v\}_{v\in[n_2]},\{\boldsymbol{x}_s\}_{s\in[n_1]},\boldsymbol{\Omega}}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > \epsilon|\boldsymbol{y}_u = y_u)\\
&\quad = \Pr_{\boldsymbol{x}_i,\boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > \epsilon|\boldsymbol{y}_u = y_u)\\
&\quad = \Pr_{\boldsymbol{x}_i,\boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > \epsilon) > 1 - r(\frac{\epsilon}{2}).
\end{aligned}$$

Since the above bound holds for all y_u such that $A \cap B$ holds, taking the expectation of the above bound with respect to $\boldsymbol{y}_u$ over the set $A \cap B$ gives

$$\Pr(|f(\boldsymbol{x}_i, \boldsymbol{y}_u) - f(\boldsymbol{x}_j, \boldsymbol{y}_u)| > \epsilon|A, B) > 1 - r(\frac{\epsilon}{2}).$$

Thus,

$$\Pr(C^c|A, B, E) < \frac{\Pr(C^c|A, B)}{1 - r(\frac{\epsilon}{2})}. \tag{9}$$

Since by hypothesis $\frac{1}{2} > \alpha > \alpha' > 0$, $p \geqslant \max(n_1^{-\frac{1}{2}+\alpha}, n_2^{-\frac{1}{2}+\alpha})$ and $k = n_2^{\alpha'}$, if n_2 is sufficiently large, the bound in line (8) is less than $\frac{1}{2}$. Then, by Bayes rule, the union bound, and Lemma D.10,

$$\begin{aligned}
\Pr(C^c|A, B) &\leqslant \frac{\Pr(C^c|A)}{\Pr(B|A)} \leqslant 2\Pr(C^c|A)\\
&= 2\Pr(\exists v \in W_u^{i,j}(\beta), |R_{\boldsymbol{uv}} - \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\delta}{4}|A)\\
&\leqslant 3(n_2 - 1)p^2 \Pr(|R_{\boldsymbol{uv}} - \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\delta}{4}|A) \qquad (10)\\
&\leqslant 3(n_2 - 1)p^2 \exp(-\frac{\delta^2}{4}\left\lfloor\frac{\beta}{2}\right\rfloor)\\
&= 3(n_2 - 1)p^2 \exp(-\frac{\delta^2}{4}\left\lfloor\frac{n_1p^2}{4}\right\rfloor)\\
&\leqslant 3(n_2 - 1)p^2 \exp(-\frac{\delta^2 n_1 p^2}{20}) \qquad (11)
\end{aligned}$$

where line (10) follows by the union bound and line (11) follows because $n_1p^2 \geqslant 16$ and $\forall x \geqslant 15$, $\left\lfloor\frac{x}{4}\right\rfloor \geqslant \frac{x}{5}$.

Since by hypothesis $\frac{1}{2} > \alpha > 0$, $p \geqslant \max(n_1^{-\frac{1}{2}+\alpha}, n_2^{-\frac{1}{2}+\alpha})$, and $n_1 \geqslant C_1 \log(n_2)^{\frac{1}{2\alpha}}$ for some constant C_1, if n_2 is sufficiently large, the bound in line (9) is eventually less than $\frac{1}{2}$. Thus, using Bayes rule and Lemma D.11,

$$\begin{aligned}
\Pr(M^c|A, B, C, E) &\leqslant \frac{\Pr(M^c|A, B, E)}{\Pr(C|A, B, E)}\\
&\leqslant 2\Pr(M^c|A, B, E)\\
&\leqslant 2\exp(-\gamma(\frac{\epsilon}{4})k). \qquad (12)
\end{aligned}$$

Putting together lines (3), (4), (5), (7), (8), (9), (11), and (12) we have

$$\begin{aligned}\Pr(D^{\epsilon}_{u,i,j}) \leqslant & 2\exp(-\frac{(n_2-1)p^2}{12}) + (n_2-1)\exp(-\frac{n_1p^2}{8}) + 2\exp(-\gamma(\frac{\epsilon}{4})k) \\ & + \frac{1}{1-r(\frac{\epsilon}{2})}[3(n_2-1)p^2\exp(-\frac{\delta^2 n_1 p^2}{20}) \\ & + \exp([1-\kappa(\frac{\epsilon}{4}) + \tau(\eta) + \log(3\frac{(n_2-1)p^2}{2})]k - k\log(k) - \tau(\eta)\frac{(n_2-1)p^2}{2})].\end{aligned}$$

□

Proof of Theorem 2. The proof follows the same steps as the proof of Theorem 1, but applies Theorem C.1 instead of Theorem B.1. □

Proof of Corollary 2. The only new term that did not appear in Corollary 2 is, ignoring constants, of the form

$$n_2^2 n_1^2 \exp(\log(n_2p^2)k - k\log(k) - n_2p^2).$$

Using $\alpha > \alpha'$ and $n_1 \leqslant C_1 n_2$, as $n_2 \longrightarrow \infty$,

$$\begin{aligned}& n_2^2 n_1^2 \exp(\log(n_2p^2)k - k\log(k) - n_2p^2) \\ & \qquad \leqslant \exp(2\log(n_2) + 2\log(n_1) + \log(n_2^{2\alpha})n_2^{\alpha'} - n_2^{\alpha'}\log(n_2)\alpha' - n_2^{2\alpha}) \\ & \qquad \leqslant \exp((2+2C_1)\log(n_2) + (2\alpha-\alpha')\log(n_2)n_2^{\alpha'} - n_2^{2\alpha}) \\ & \qquad \longrightarrow 0\end{aligned}$$

□

D Technical Lemmas

We separate the lemmas into three sections: lemmas for both the continuous and discrete rating settings, lemmas for the continuous rating setting, and lemmas for the discrete rating setting.

D.1 Lemmas Common to the Continuous Rating Setting and the Discrete Rating Setting

Lemma D.1 establishes that for a user $u \in [n_2]$ and distinct items $i, j \in [n_1]$, with high probability there are many other users that have rated items i and j and many items in common with user u. It is similar to Lemma 1 from Lee et al. (2016).

Lemma D.1. *Fix $u \in [n_2]$, $i \neq j \in [n_1]$, and let $\lambda > 0$ and $2 \leqslant \beta \leqslant \frac{n_1p^2}{2}$. Let $W_u^{i,j}(\beta) = \{v \in [n_2] : |N(u,v)| \geqslant \beta, (i,v), (j,v) \in \Omega\}$. Then,*

$$\begin{aligned}& \Pr_{\mathbf{\Omega}}(|W_u^{i,j}(\beta)| \notin [(1-\lambda)(n_2-1)p^2, (1-\lambda)(n_2-1)p^2]) \\ & \qquad \leqslant 2\exp(-\frac{\lambda^2(n_2-1)p^2}{3}) + (n_2-1)\exp(-\frac{n_1p^2}{8}).\end{aligned}$$

Proof. Define the following binary variables for all $v \in [n_2]\backslash\{u\}$. $E_v = 1$ if $|N(u,v)| \geqslant \beta$ and 0 otherwise, $F_v = 1$ if $(i,v) \in \mathbf{\Omega}$ and 0 otherwise, and $G_v = 1$ if $(j,v) \in \mathbf{\Omega}$ and 0 otherwise. Observe that $|W_u^{i,j}(\beta)| = \sum_{v \neq u} E_v F_v G_v$. Fix $0 \leqslant a < b \leqslant n_2 - 1$. Observe that if $\sum_{v\neq u} F_v G_v \in [a,b]$ and $\sum_{v\neq u} E_v = n_2 - 1$, then $|W_u^{i,j}(\beta)| \in [a,b]$. Thus, the contrapositive implies that for any $0 \leqslant a < b \leqslant n_2 - 1$,

$$\begin{aligned}\Pr_{\mathbf{\Omega}}(|W_u^{i,j}(\beta)| \notin [a,b]) &\leqslant \Pr_{\mathbf{\Omega}}(\sum_{v\neq u} F_v G_v \notin [a,b] \cup \sum_{v\neq u} E_v < n_2 - 1)\\ &\leqslant \Pr_{\mathbf{\Omega}}(\sum_{v\neq u} F_v G_v \notin [a,b]) + \Pr_{\mathbf{\Omega}}(\sum_{v\neq u} E_v < n_2 - 1).\end{aligned}$$

$\sum_{v\neq u} F_v G_v$ is a binomial random variable with parameters $n_2 - 1$ and p^2. Letting $a = (1-\lambda)(n_2 - 1)p^2$ and $b = (1+\lambda)(n_2-1)p^2$, Chernoff's multiplicative bound (Proposition G.2) yields that

$$\Pr_{\mathbf{\Omega}}(\sum_{v\neq u} F_v G_v \notin [(1-\lambda)(n_2-1)p^2, (1+\lambda)(n_2-1)p^2]) \leqslant 2\exp(-\frac{\lambda^2(n_2-1)p^2}{3}).$$

Since $N(u,v)$ is binomial with parameters n_1 and p^2, by Chernoff's multiplicative bound (Proposition G.2),

$$\begin{aligned}\Pr_{\mathbf{\Omega}}(E_v = 0) &= \Pr_{\mathbf{\Omega}}(N(u,v) \leqslant \beta)\\ &\leqslant \Pr_{\mathbf{\Omega}}(N(u,v) \leqslant \frac{n_1 p^2}{2})\\ &\leqslant \exp(-\frac{n_1 p^2}{8}).\end{aligned}$$

Then, by the union bound,

$$\begin{aligned}\Pr_{\mathbf{\Omega}}(\sum_{v\neq u} E_v < n_2 - 1) &= \Pr_{\mathbf{\Omega}}(\exists v \in [n_2]\backslash\{u\} : E_v = 0)\\ &\leqslant (n_2 - 1)\exp(-\frac{n_1 p^2}{8}).\end{aligned}$$

□

To convert the pairwise comparisons to a ranking, we use the Copeland ranking procedure (Algorithm 3 in the main document). Lemma D.2 establishes that if the output of the Pairwise-Rank algorithm is such that for all $i, j \in [n_1]$ and $u \in [n_2]$, $D^{\epsilon}_{u,i,j}$ does not occur, then applying the Copeland ranking procedure to A (as defined in Multi-Rank) yields a $\hat{\sigma}$ such that $\text{dis}_{2\epsilon}(\hat{\sigma}, H) = 0$.

Lemma D.2. *Let $\epsilon > 0$, $u \in [n_2]$, A as defined in Multi-Rank (Algorithm 1), and $\hat{\sigma}_u$ = Copeland($A_{u,:,:}$). If for all $i \neq j \in [n_1]$ $f(x_i, y_u) > f(x_j, y_u) + \epsilon$ implies that $A_{u,i,j} = 1$, then for all $i \neq j \in [n_1]$ $h_u(x_i, y_u) > h_u(x_j, y_u)$ and $f(x_i, y_u) > f(x_j, y_u) + 2\epsilon$ implies that $\hat{\sigma}_u(i) > \hat{\sigma}_u(j)$.*

Proof. Let $i \neq j \in [n_1]$ such that $h_u(x_i, y_u) > h_u(x_j, y_u)$ and $f(x_i, y_u) > f(x_j, y_u) + 2\epsilon$. Let $l \in [n_1]$ such that $l \neq i$ and $l \neq j$. We claim that if $A_{u,i,l} = 0$, then $A_{u,j,l} = 0$. If $A_{u,i,l} = 0$, then by the hypothesis $f(x_i, y_u) \leqslant f(x_l, y_u) + \epsilon$. Then,

$$f(x_j, y_u) + 2\epsilon < f(x_i, y_u) \leqslant f(x_l, y_u) + \epsilon$$

so that $f(x_j, y_u) + \epsilon < f(x_l, y_u)$. Then, by the hypothesis, $A_{u,j,l} = 0$, establishing the claim.

The contrapositive of the claim is that if $A_{u,j,l} = 1$, then $A_{u,i,l} = 1$. Then,

$$I_j = \sum_{l=1, l\neq j}^{n_1} A_{u,j,l} = \sum_{l=1, l\notin\{j,i\}}^{n_1} A_{u,j,l} \leqslant \sum_{l=1, l\notin\{j,i\}}^{n_1} A_{u,i,l} = I_i - 1 < I_i$$

so that $\widehat{\sigma}_u(i) > \widehat{\sigma}_u(j)$. □

Recall the definition of our problem-specific constants: $\tau(\epsilon) = \inf_{y_0 \in \mathcal{Y}} \Pr_{\boldsymbol{y}_u}(d_{\mathcal{Y}}(y_0, \boldsymbol{y}_u) \leqslant \epsilon)$, $\kappa(\epsilon) = \inf_{y_0 \in \mathcal{Y}} \Pr_{\boldsymbol{y}_u}(d_{\mathcal{Y}}(y_0, \boldsymbol{y}_u) > \epsilon)$, and $\gamma(\epsilon) = \inf_{z \in [-N,N]} \mathcal{P}_{\{\boldsymbol{a}_{u,l}\}_{l \in [L-1]}}(\exists l \in [L-1] : d_{\mathbb{R}}(z, \boldsymbol{a}_{u,l}) \leqslant \epsilon)$. Lemma D.3 establishes that under our assumptions, for all $\epsilon > 0$, $\tau(\epsilon) > 0$, $\kappa(\epsilon) < 1$, and $\gamma(\epsilon) > 0$.

Lemma D.3. *If there exists $\epsilon > 0$ such that $\tau(\epsilon) = 0$, or $\kappa(\epsilon) = 1$, then there exists a point $z \in \mathcal{Y}$ such that $\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z)) = 0$. Similarly, if there exists $\epsilon > 0$ such that $\gamma(\epsilon) = 0$, then there exists $z \in [-N, N]$ such that $\mathcal{P}_l(B_\epsilon(z)) = 0$ for all $l \in [L-1]$.*

Proof. Let $\epsilon > 0$ and suppose $\tau(\epsilon) = 0$. Then, there exists a sequence of points $z_1, z_2, \ldots \in \mathcal{Y}$ such that for every n, $\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z_n)) \leqslant \frac{1}{n}$. Since $\mathcal{Y}$ is compact by assumption, there exists a convergent subsequence $z_{i_1}, z_{i_2}, \ldots$ to z.

We claim that for all $z' \in \mathcal{Y}$, there exists a sufficiently large N such that $z' \in B_\epsilon(z_{i_N})$ if and only if $z' \in B_\epsilon(z)$. Fix $z' \in B_\epsilon(z)$. Since $B_\epsilon(z)$ is open, there exists $\delta > 0$ such that $d_{\mathcal{Y}}(z, z') < \delta < \epsilon$. Let N large enough such that $d_{\mathcal{Y}}(z, z_{i_N}) \leqslant \epsilon - \delta$. Then, by the triangle inequality,

$$d(z', z_{i_N}) \leqslant d(z', z) + d(z_{i_N}, z) \leqslant \delta + \epsilon - \delta = \epsilon$$

so that $z' \in B_\epsilon(z_{i_N})$. A similar argument shows the other direction of the claim. Since a probability space has finite measure, by the dominated convergence theorem,

$$\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z)) = \lim_{n \longrightarrow \infty} \mathcal{P}_{\mathcal{Y}}(B_\epsilon(z_{i_n})) \leqslant \lim_{n \longrightarrow \infty} \frac{1}{i_n} = 0.$$

Next, suppose $\kappa(\epsilon) = 1$. Then, there exists a sequence of points $z_1, z_2, \ldots \in \mathcal{Y}$ such that for every n, $\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z_n)^c) \geqslant 1 - \frac{1}{n}$. Then, for every n, $\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z_n)) \leqslant \frac{1}{n}$ A similar argument from the $\tau(\cdot)$ case using the dominated convergence theorem shows that $\mathcal{P}_{\mathcal{Y}}(B_\epsilon(z)) = 0$.

Since $[-N, N]$ is compact and γ has a similar definition to τ, the result for $\gamma(\cdot)$ follows by an argument similar to the one used for the $\tau(\cdot)$ case. □

D.2 Lemmas for Continuous Rating Setting

Lemma D.4 uses the notion of r-discerning to relate the distance between points in $\mathcal{Y}$ and to a lower bound on $\rho(y_u, y_v)$.

Lemma D.4. *Let r be a positive nondecreasing function. If $y_u \in \mathcal{Y}$ is r-discerning, then for any $\epsilon > 0$, if $y_v \in B_\epsilon(y_u)$, then $\rho(y_u, y_v) > 1 - r(\epsilon)$.*

Proof. Suppose that $d(y_u, y_v) \leqslant \epsilon$. Suppose that $\boldsymbol{x}_i = x_i$ and $\boldsymbol{x}_j = x_j$ such that $|f(x_i, y_u) - f(x_j, y_u)| > 2\epsilon$ and without loss of generality suppose that $h_u(x_i, y_u) \geqslant h_u(x_j, y_u)$. Then, since f is Lipschitz,

$$f(x_i, y_v) \geqslant f(x_i, y_u) - \epsilon > f(x_j, y_u) + \epsilon \geqslant f(x_j, y_v).$$

Hence, $h_v(x_i, y_v) \geqslant h_v(x_j, y_v)$. Thus,

$$\rho(y_u, y_v) \geqslant \mathrm{Pr}_{\boldsymbol{x}_i,\boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > 2\epsilon) > 1 - r(\epsilon),$$

where the last inequality follows from the hypothesis that y_u is r-discerning. Thus, we conclude the result. □

Lemma D.5 establishes that if $S \subset [n_2]\backslash\{u\}$ is a large enough set, then with high probability there is at least one element $\boldsymbol{y}_v$ in S that tends to agree with $\boldsymbol{y}_u$.

Lemma D.5. *Let r be a positive non-decreasing function and suppose that almost every $y \in \mathcal{Y}$ is r-discerning. Let $S \subset [n_2]\backslash\{u\}$. Then, $\forall \epsilon > 0$,*

$$\mathrm{Pr}_{\boldsymbol{y}_v,\boldsymbol{y}_u}(\max_{v\in[S]} \rho(\boldsymbol{y}_v, \boldsymbol{y}_u) \leqslant 1 - r(\epsilon)) \leqslant [1 - \tau(\epsilon)]^{|S|}.$$

Proof. Fix $\boldsymbol{y}_u = y_u \in \mathcal{Y}$ that is r-discerning. By Lemma D.4, if $\boldsymbol{y}_v = y_v$ is such that $d(y_u, y_v) \leqslant \epsilon$, then $\rho(y_u, y_v) > 1 - r(\epsilon)$. Hence,

$$\mathrm{Pr}_{\boldsymbol{y}_v}(d(y_u, \boldsymbol{y}_v)) \leqslant \epsilon) \leqslant \mathrm{Pr}_{\boldsymbol{y}_v}(\rho(y_u, \boldsymbol{y}_v) > 1 - r(\epsilon)).$$

Then,

$$\mathrm{Pr}_{\boldsymbol{y}_v}(\rho(y_u, \boldsymbol{y}_v) \leqslant 1 - r(\epsilon)) \leqslant \mathrm{Pr}_{\boldsymbol{y}_v}(d(y_u, \boldsymbol{y}_v)) > \epsilon) = 1 - \mathrm{Pr}_{\boldsymbol{y}_v}(d(y_u, \boldsymbol{y}_v)) \leqslant \epsilon) \leqslant 1 - \tau(\epsilon).$$

The RHS does not depend on y_u, and $\boldsymbol{y}_v, \boldsymbol{y}_u$ are independent and almost every $y \in \mathcal{Y}$ is r-discerning, so we can take the expectation with respect to $\boldsymbol{y}_u$ to obtain

$$\mathrm{Pr}_{\boldsymbol{y}_v,\boldsymbol{y}_u}(\rho(\boldsymbol{y}_v, \boldsymbol{y}_u) \leqslant 1 - r(\epsilon)) \leqslant 1 - \tau(\epsilon). \tag{13}$$

Finally,

$$\begin{aligned}\mathrm{Pr}_{\{\boldsymbol{y}_v\}_{v\in S},\boldsymbol{y}_u}(\max_{v\in[S]} \rho(\boldsymbol{y}_v, \boldsymbol{y}_u) \leqslant 1 - r(\epsilon)) &= \mathrm{Pr}_{\boldsymbol{y}_v,\boldsymbol{y}_u}(\rho(\boldsymbol{y}_v, \boldsymbol{y}_u) \leqslant 1 - r(\epsilon))^{|S|} \\ &\leqslant [1 - \tau(\epsilon)]^{|S|},\end{aligned}$$

where the first equality follows from the independence of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2}$ and the inequality follows from line (13). □

Lemma D.6 establishes that $R_{\boldsymbol{u},\boldsymbol{v}}$ concentrates around $\rho(\boldsymbol{y}_u, \boldsymbol{y}_v)$.

Lemma D.6. *Let $u \neq v \in [n_2]$, $i \neq j \in [n_1]$, $\eta > 0$, $\beta \geqslant 2$, and $W_u^{i,j}(\beta)$ be defined as in Lemma D.1. Then,*

$$\mathrm{Pr}(|R_{\boldsymbol{u},\boldsymbol{v}} - \rho(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\eta}{4}|v \in W_u^{i,j}(\beta)) \leqslant 2\exp(-\frac{\eta^2}{4}\left\lfloor\frac{\beta}{2}\right\rfloor).$$

Proof. Fix $\boldsymbol{y}_u = y_u$ and $\boldsymbol{y}_v = y_v$. Recall that if $I(u, v) \neq \varnothing$, then

$$R_{u,v} = \frac{1}{|I(u,v)|} \sum_{(s,t)\in I(u,v)} \mathbf{1}\{(h_u(\boldsymbol{x}_s, y_u) - h_u(\boldsymbol{x}_t, y_u))(h_v(\boldsymbol{x}_s, y_v) - h_v(\boldsymbol{x}_t, y_v)) \geqslant 0\}.$$

Since $I(u, v)$ consists of pairs of indices that do not overlap, conditioned on $\boldsymbol{y}_u = y_u, \boldsymbol{y}_v = y_v$, and any nonempty $I(u, v)$, $\{\mathbf{1}\{(h_u(\boldsymbol{x}_s, y_u) - h_u(\boldsymbol{x}_t, y_u))(h_v(\boldsymbol{x}_s, y_v) - h_v(\boldsymbol{x}_t, y_v)) \geqslant 0\} : (s, t) \in I(u, v)\}$ is a set of independent random variables. Further, each has mean $\rho(y_u, y_v)$. Thus, by Chernoff's bound (Proposition G.1),

$$\Pr(|R_{u,v} - \rho(y_u, y_v)| > \frac{\eta}{4} | \boldsymbol{y}_u = y_u, \boldsymbol{y}_v = y_v, I(u, v)) \leqslant \exp(-\frac{\eta^2}{2}|I(u, v)|)$$

When $v \in v \in W_u^{i,j}(\beta)$, $|I(u, v)| \geqslant \left\lfloor \frac{\beta}{2} \right\rfloor$. Since the above bound holds for all y_u, y_v, it follows that

$$\Pr(|R_{\boldsymbol{u},\boldsymbol{v}} - \rho(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\eta}{4} | v \in W_u^{i,j}(\beta)) \leqslant 2\exp(-\frac{\eta^2}{4}\left\lfloor \frac{\beta}{2} \right\rfloor).$$

□

Lemma D.7 establishes that conditional on A, B, C (defined in the proof of Theorem B.1), the event $D^{\epsilon}_{u,i,j}$ does not occur with probability 1.

Lemma D.7. *Under the setting described in Theorem B.1, let $u \in [n_2]$ and $i \neq j \in [n_1]$. Then, $\Pr(D^{\epsilon}_{u,i,j}|A, B, C) = 0$.*

Proof. Define the events

$$\begin{aligned} E_1 &= \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) + \epsilon < f(\boldsymbol{x}_j, \boldsymbol{y}_u)\} \\ E_2 &= \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) > f(\boldsymbol{x}_j, \boldsymbol{y}_u) + \epsilon\} \end{aligned}$$

By the union bound and law of total probability,

$$\begin{aligned} \Pr(D^{\epsilon}_{u,i,j}|A, B, C) &\leqslant \Pr(\mathrm{PR}(u, i, j, \beta, k) = 1 \cap E_1 | A, B, C) \\ &\quad + \Pr(\mathrm{PR}(u, i, j, \beta, k) = 0 \cap E_2 | A, B, C) \\ &\leqslant \Pr(\mathrm{PR}(u, i, j, \beta, k) = 1 | A, B, C, E_1) \\ &\quad + \Pr(\mathrm{PR}(u, i, j, \beta, k) = 0 | A, B, C, E_2). \end{aligned}$$

The argument for bounding each of these is similar and, thus, we bound the term $\Pr(\mathrm{PR}(u, i, j, \beta, k) = 1 | A, B, C, E_1)$.

Fix $\{\boldsymbol{y}_v = y_v\}_{v \in [n_2]}$ r-discerning and $(\frac{\epsilon}{2}, \delta)$-discriminative, $\{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}$, and $\boldsymbol{\Omega} = \Omega$ such that the event $A \cap B \cap C \cap E_1$ occurs. We claim that Pairwise-Rank puts $V = \{v\}$ (see Algorithm 2 for definition of V) such that $y_v \in B_{\frac{\epsilon}{2}}(y_u)$. On the event B, there is $v \in W_u^{i,j}(\beta)$ with $\rho(y_u, y_v) \geqslant 1 - \frac{\delta}{2}$. Since y_u is $(\frac{\epsilon}{2}, \delta)$-discriminative, it follows that $y_v \in B_{\frac{\epsilon}{2}}(y_u)$. Suppose that $w \in W_u^{i,j}(\beta)$ such that $y_w \in B_{\frac{\epsilon}{2}}(y_u)^c$. Since y_u is $(\frac{\epsilon}{2}, \delta)$-discriminative, $\rho(y_w, y_u) < 1 - \delta$. Then,

$$\begin{aligned} R_{w,u} &\leqslant \rho(y_w, y_u) + \frac{\delta}{4} && (14) \\ &< 1 - \frac{3}{4}\delta \\ &\leqslant \rho(y_u, y_v) - \frac{\delta}{4} \\ &\leqslant R_{u,v} && (15) \end{aligned}$$

where lines (14) and (15) follow by event C and $v, w \in W_u^{i,j}(\beta)$. Thus, the claim follows. Conditional on E_1, we have that $f(x_i, y_u) + \epsilon < f(x_j, y_u)$. Then, using the Lipschitzness of f,

$$f(x_i, y_v) \leqslant f(x_i, y_u) + \frac{\epsilon}{2} < f(x_j, y_u) - \frac{\epsilon}{2} \leqslant f(x_j, y_v).$$

Since g_v is strictly increasing by hypothesis, $h_v(x_i, y_v) < h_v(x_j, y_v)$. Thus, Pairwise-Rank with $k = 1$ outputs 0. Consequently,

$$\Pr(\mathrm{PR}(u, i, j, \beta, k) = 1 | A, B, C, E_1, \{\boldsymbol{y}_v = y_v\}_{v \in [n_2]}\{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}, \boldsymbol{\Omega} = \Omega) = 0$$

Since almost every $y \in \mathcal{Y}$ is r-discerning and $(\frac{\epsilon}{2}, \delta)$-discriminative, taking the expectation wrt $\{\boldsymbol{y}_v\}_{v \in [n_2]}$, $\{\boldsymbol{x}_s\}_{s \in [n_1]}$, $\boldsymbol{\Omega}$ on the set $A \cap B \cap C \cap E_1$ of the last equality gives the result. □

D.3 Lemmas for Discrete Rating Setting

Lemma D.8 is the analogoue of Lemma D.4 for the discrete case. The proof is very similar.

Lemma D.8. *Let r be a positive non-decreasing function. If $y_u \in \mathcal{Y}$ is r-discerning, then for any $\epsilon > 0$, if $y_v \in B_\epsilon(y_u)$, then $\rho'(y_u, y_v) > 1 - r(\epsilon)$.*

Proof. Suppose y_v is such that $d(y_u, y_v) \leqslant \epsilon$. We claim that under this assumption

$$\rho'(y_u, y_v) \geqslant \Pr_{\boldsymbol{x}_i, \boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > 2\epsilon). \tag{16}$$

Fix $\boldsymbol{g}_u = g_u$ and $\boldsymbol{g}_v = g_v$, and $\boldsymbol{x}_i = x_i$ and $\boldsymbol{x}_j = x_j$ such that $|f(x_i, y_u) - f(x_j, y_u)| > 2\epsilon$. Without loss of generality, suppose that $h_u(x_i, y_u) \geqslant h_u(x_j, y_u)$. Then, since f is Lipschitz,

$$f(x_i, y_v) \geqslant f(x_i, y_u) - \epsilon > f(x_j, y_u) + \epsilon \geqslant f(x_j, y_v).$$

Hence, $h_v(x_i, y_v) \geqslant h_v(x_j, y_v)$, establishing that

$$\begin{aligned} &\rho'(y_u, y_v | \boldsymbol{g}_u = g_u, \boldsymbol{g}_v = g_v) \\ &\quad = \Pr_{\boldsymbol{x}_i, \boldsymbol{x}_j}([g_u(f(\boldsymbol{x}_i, y_u)) - g_u(f(\boldsymbol{x}_j, y_u))][g_v(f(\boldsymbol{x}_i, y_v)) - g_u(f(\boldsymbol{x}_j, y_v))] \geqslant 0) \\ &\quad \geqslant \Pr_{\boldsymbol{x}_i, \boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > 2\epsilon). \end{aligned} \tag{17}$$

Since $\{\boldsymbol{g}_u, \boldsymbol{g}_v, \boldsymbol{x}_i, \boldsymbol{x}_j\}$ are independent, taking the expectation with respect to $\boldsymbol{g}_u$ and $\boldsymbol{g}_v$ in line (17) establishes line (16). Thus,

$$\rho'(y_u, y_v) \geqslant \Pr_{\boldsymbol{x}_i, \boldsymbol{x}_j}(|f(\boldsymbol{x}_i, y_u) - f(\boldsymbol{x}_j, y_u)| > 2\epsilon) > 1 - r(\epsilon),$$

where the last inequality follows from the hypothesis that y_u is r-discerning. □

Lemma D.9 is the analogoue of Lemma D.5 for the discrete case.

Lemma D.9. *Let $\epsilon, \delta > 0$. Let r be a positive nondecreasing function such that $r(\epsilon) \geqslant \delta$ and $r(\eta) < \delta$ for some $\eta > 0$. Suppose that almost every $y \in \mathcal{Y}$ is (ϵ, δ)-discriminative and r-discerning. Let $R_2 \geqslant R_1 \geqslant 0$ be constants. Then, for any $S \subset [n_2]$ depending on $\boldsymbol{\Omega}$ and $k \leqslant R_1$,*

$$\begin{aligned} &\Pr_{\boldsymbol{y}_v, \boldsymbol{y}_u}(\max^{(k)}_{v \in [S]} \rho'(\boldsymbol{y}_v, \boldsymbol{y}_u) \leqslant 1 - r(\eta) \,|R_1 \leqslant |S| \leqslant R_2) \\ &\quad \leqslant \exp((1 - \kappa(\epsilon) + \tau(\eta) + \log(R_2))k - k\log(k) - \tau(\eta)R_1) | R_1 \leqslant |S| \leqslant R_2). \end{aligned}$$

Proof. Let $C_\eta = \Pr_{\boldsymbol{y}_v,\boldsymbol{y}_u}(\rho'(\boldsymbol{y}_v,\boldsymbol{y}_u) \leqslant 1 - r(\eta))$.

Claim: $C_\eta \leqslant 1 - \tau(\eta)$.

Fix $\boldsymbol{y}_u = y_u \in \mathcal{Y}$ r-discerning. By Lemma D.8, if $\boldsymbol{y}_v = y_v$ is such that $d(y_u, y_v) \leqslant \epsilon$, then $\rho'(y_u, y_v) > 1 - r(\epsilon)$. Hence,

$$\Pr_{\boldsymbol{y}_v}(d(y_u,\boldsymbol{y}_v)) \leqslant \epsilon) \leqslant \Pr_{\boldsymbol{y}_v}(\rho'(y_u,\boldsymbol{y}_v) > 1 - r(\epsilon)).$$

Then,

$$\Pr_{\boldsymbol{y}_v}(\rho'(y_u,\boldsymbol{y}_v) \leqslant 1 - r(\epsilon)) \leqslant \Pr_{\boldsymbol{y}_v}(d(y_u,\boldsymbol{y}_v)) > \epsilon) = 1 - \Pr_{\boldsymbol{y}_v}(d(y_u,\boldsymbol{y}_v)) \leqslant \epsilon) \leqslant 1 - \tau(\epsilon),$$

where the last inequality follows by the definition of $\tau(\cdot)$. The RHS does not depend on y_u, and $\boldsymbol{y}_v, \boldsymbol{y}_u$ are independent, so we can take the expectation with respect to $\boldsymbol{y}_u$ to establish the claim.

Claim: $1 - C_\eta \leqslant 1 - \kappa(\epsilon)$.

Since almost every $y \in \mathcal{Y}$ is (ϵ, δ)-discriminative and $r(\eta) < \delta$, $\mathcal{Y}$ is almost-everywhere $(\epsilon, r(\eta))$-discriminative. Fix $\boldsymbol{y}_u = y_u$ such that y_u is $(\epsilon, r(\eta))$-discriminative. Then, $\forall y_v \in \mathcal{Y}$, $\rho'(y_u, y_v) > 1 - r(\eta)$ implies that $d_{\mathcal{Y}}(y_u, y_v) \leqslant \epsilon$. Thus,

$$\begin{aligned}
\Pr_{\boldsymbol{y}_v}(\rho'(y_u,\boldsymbol{y}_v) > 1 - r(\eta)) &\leqslant \Pr_{\boldsymbol{y}_v}(d_{\mathcal{Y}}(y_u,\boldsymbol{y}_v \leqslant \epsilon) \\
&= 1 - \Pr_{\boldsymbol{y}_v}(d_{\mathcal{Y}}(y_u,\boldsymbol{y}_v) > \epsilon) \\
&\leqslant 1 - \kappa(\epsilon).
\end{aligned}$$

Since the RHS does not depend on y_u, and $\boldsymbol{y}_u$ and $\boldsymbol{y}_v$ are independent, we can take the expectation with respect to $\boldsymbol{y}_u$ to establish the claim.

Main Probability Bound: Fix $\boldsymbol{\Omega} = \Omega$ such that $R_1 \leqslant |S| \leqslant R_2$.

$$\begin{aligned}
\Pr_{\boldsymbol{y}_v,\boldsymbol{y}_u}(\max^{(k)}_{v\in[S]}\rho'(\boldsymbol{y}_v,\boldsymbol{y}_u) \leqslant & 1 - r(\eta)|\boldsymbol{\Omega} = \Omega) \\
&= \sum_{l=0}^{k-1} \binom{|S|}{l} C_\eta^{|S|-l}(1 - C_\eta)^l \\
&\leqslant k \max_{l\in\{0,\dots,k-1\}} \binom{|S|}{l} C_\eta^{|S|-l}(1 - C_\eta)^l \\
&\leqslant k \max_{l\in[k-1]\cup\{0\}} \binom{|S|}{l} (1 - \tau(\eta))^{|S|-l}(1 - \kappa(\epsilon))^l \\
&\leqslant k \max_{l\in\{0,\dots,k-1\}} (\frac{|S|e}{l})^l (1 - \tau(\eta))^{|S|-l}(1 - \kappa(\epsilon))^l & (18) \\
&\leqslant k \max_{l\in\{0,\dots,k-1\}} \exp(l + l\log(\frac{|S|}{l}) - \tau(\eta)[|S| - l] - \kappa(\epsilon)l) & (19) \\
&= k \max_{l\in\{0,\dots,k-1\}} \exp([1 - \kappa(\epsilon) + \tau(\eta)]l + l\log(\frac{|S|}{l}) - \tau(\eta)|S|)) \\
&\leqslant k\exp([1 - \kappa(\epsilon) + \tau(\eta)]k + k\log(\frac{|S|}{k}) - \tau(\eta)|S|)) & (20) \\
&= \exp([1 - \kappa(\epsilon) + \tau(\eta) + \log(|S|)]k - k\log(k) - \tau(\eta)|S|)) \\
&\leqslant \exp([1 - \kappa(\epsilon) + \tau(\eta) + \log(R_1)]k - k\log(k) - \tau(\eta)R_2))
\end{aligned}$$

where line (18) follows from the the inequality $\binom{n}{k} \leqslant (\frac{ne}{k})^k$, line (19) follows from the inequality $(1-x) \leqslant \exp(-x)$, and line (20) follows since $|S| \geqslant k$ and $1-\kappa(\epsilon) > 0$ by Lemma D.3. Finally, we can take the expectation with respect to $\boldsymbol{\Omega} = \Omega$ over the set $R_1 \leqslant |S| \leqslant R_2$ to conclude the result. □

Lemma D.10 is the analogoue of Lemma D.6 for the discrete case.

Lemma D.10. *Consider the discrete ratings setting. Let $u \neq v \in [n_2]$, $i \neq j \in [n_1]$, $\eta > 0$, $\beta \geqslant 2$, and $W_u^{i,j}(\beta)$ be defined as in Lemma D.1. Then,*

$$\Pr(|R_{\boldsymbol{u},\boldsymbol{v}} - \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\eta}{4} | v \in W_u^{i,j}(\beta)) \leqslant 2\exp(-\frac{\eta^2}{4}\left\lfloor \frac{\beta}{2} \right\rfloor).$$

Proof. Fix $\boldsymbol{y}_u = y_u$, $\boldsymbol{y}_v = y_v$, and $\boldsymbol{g}_u = g_u, \boldsymbol{g}_v = g_v$. Recall that if $I(u,v) \neq \varnothing$, then

$$R_{u,v} = \frac{1}{|I(u,v)|} \sum_{(s,t) \in I(u,v)} \mathbf{1}\{[h_u(\boldsymbol{x}_s, y_u) - h_u(\boldsymbol{x}_t, y_u)][h_v(\boldsymbol{x}_s, y_v) - h_v(\boldsymbol{x}_t, y_v)] \geqslant 0\}.$$

Since $I(u,v)$ consists of pairs of indices that do not overlap, conditioned on $\boldsymbol{y}_v = y_v$, $\boldsymbol{y}_u = y_u$, $\boldsymbol{g}_u = g_u$, $\boldsymbol{g}_v = g_v$ and any nonempty $I(u,v)$,

$$\{\mathbf{1}\{(g_u(f(\boldsymbol{x}_s, y_u)) - g_u(f(\boldsymbol{x}_t, y_u)))(g_v(f(\boldsymbol{x}_s, y_v)) - g_v(f(\boldsymbol{x}_t, y_v))) \geqslant 0\} : (s,t) \in I(u,v)\}$$

is a set of independent random variables. Further, each has mean $\rho'(y_u, y_v | \boldsymbol{g}_u = g_u, \boldsymbol{g}_v = g_v)$. Thus, by Chernoff's bound (Proposition G.1),

$$\begin{aligned} &\Pr(|R_{u,v} - \rho'(y_u, y_v | \boldsymbol{g}_u = g_u, \boldsymbol{g}_v = g_v)| > \frac{\eta}{4} | \boldsymbol{y}_u = y_u, \boldsymbol{y}_v = y_v, \boldsymbol{g}_u = g_u, \boldsymbol{g}_v = g_v, I(u,v)) \\ &\quad \leqslant \exp(-\frac{\eta^2}{2}|I(u,v)|) \end{aligned}$$

When $v \in v \in W_u^{i,j}(\beta)$, $|I(u,v)| \geqslant \left\lfloor \frac{\beta}{2} \right\rfloor$. Since the above bound holds for all $\boldsymbol{y}_u, \boldsymbol{y}_v, \boldsymbol{g}_u \boldsymbol{g}_v$, it follows that

$$\Pr(|R_{\boldsymbol{u},\boldsymbol{v}} - \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v)| > \frac{\eta}{4} | v \in W_u^{i,j}(\beta)) \leqslant 2\exp(-\frac{\eta^2}{4}\left\lfloor \frac{\beta}{2} \right\rfloor).$$

□

Lemma D.11. *Let $\epsilon, \delta > 0$, $\frac{1}{2} > \alpha > \alpha' > 0$, and r be a positive nondecreasing function such that $r(\frac{\epsilon}{4}) \geqslant \delta$ and $r(\eta) < \frac{\delta}{2}$ for some $\eta > 0$. Suppose that almost every $y \in \mathcal{Y}$ is r-discerning and $(\frac{\epsilon}{4}, \delta)$-discriminative. Fix $u \in [n_2]$, $i \neq j \in [n_1]$, and $k \leqslant \frac{(n_2-1)p^2}{2}$. As in the proof of Theorem C.1, define*

$$\begin{aligned} &A = \{|W_u^{i,j}(\beta)| \in [\frac{(n_2-1)p^2}{2}, \frac{3(n_2-1)p^2}{2}]\}, \\ &B = \{\max^{(k)}_{v \in W_u^{i,j}(\beta)} \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v) \geqslant 1 - \frac{\delta}{2}\}, \\ &E = \{|f(\boldsymbol{x}_i, \boldsymbol{y}_u) - f(\boldsymbol{x}_j, \boldsymbol{y}_u)| > \epsilon\} \\ &M = \{\exists v \in W_u^{i,j}(\beta) \text{ s.t. } \rho'(\boldsymbol{y}_u, \boldsymbol{y}_v) \geqslant 1 - \frac{\delta}{2} \text{ and } \exists l \in [L-1] \text{ s.t. } \boldsymbol{a}_{v,l} \in (f(\boldsymbol{x}_j, \boldsymbol{y}_v), f(\boldsymbol{x}_i, \boldsymbol{y}_v))\}. \end{aligned}$$

Then,

$$\Pr(M^c|A,B,E) \leqslant \exp(-\gamma(\frac{\epsilon}{4})k).$$

Proof. Fix $\{\boldsymbol{y}_v = y_v\}_{v\in[n_2]}$ r-discerning and $(\frac{\epsilon}{4},\delta)$-discriminative, $\boldsymbol{\Omega} = \Omega$, and $\{\boldsymbol{x}_s = x_s\}_{s\in[n_1]}$ such that $A \cap B \cap E$ holds. Let $R = \{v \in [n_2]\backslash\{u\} : v \in W_u^{i,j}(\beta) \text{ and } \rho'(y_u, y_v) \geqslant 1 - \frac{\delta}{2}\}$. Events A and B imply that $|R| \geqslant k$. Since y_u is $(\frac{\epsilon}{4},\delta)$-discriminative and for all $v \in R$, $\rho'(y_u, y_v) \geqslant 1 - \frac{\delta}{2}$, it follows that for all $v \in R$, $y_v \in B_{\frac{\epsilon}{4}}(y_u)$.

By E, $|f(x_i, y_u) - f(x_j, y_u)| > \epsilon$. Suppose that $f(x_i, y_u) > f(x_j, y_u) + \epsilon$ (the other case is similar). Then, by Lipschitzness of f, for all $v \in R$

$$f(x_j, y_v) \leqslant f(x_j, y_u) + \frac{\epsilon}{4} < f(x_i, y_u) - \frac{3}{4}\epsilon \leqslant f(x_i, y_v) - \frac{\epsilon}{2}.$$

Thus, for all $v \in R$, $(f(x_j, y_v), f(x_i, y_v))$ is an open interval of length at least $\frac{\epsilon}{2}$. Fix $v' \in [n_2]\backslash\{u\}$. Since R is a finite set, the following is well-defined:

$$I := \mathop{\arg\min}_{J \in \{(f(x_j,y_v), f(x_i,y_v)) : v \in R\}} \Pr_{\{\boldsymbol{a}_{v',l}\}_{l\in[L-1]}}(\exists l \in [L-1] \text{ s.t. } a_{v',l} \in J). \tag{21}$$

Then,

$$\begin{aligned}
&\Pr_{\{\boldsymbol{a}_{v,l}\}}(\forall v \in R, \forall l \in [L-1], \boldsymbol{a}_{v,l} \notin (f(x_j, y_v), f(x_i, y_v))|\{\boldsymbol{y}_v = y_v\}_{v\in[n_2]}, \boldsymbol{\Omega} = \Omega, \{\boldsymbol{x}_s = x_s\}_{s\in[n_2]}) \\
&\quad = \Pr_{\{\boldsymbol{a}_{v,l}\}}(\forall v \in R, \forall l \in [L-1], \boldsymbol{a}_{v,l} \notin (f(x_j, y_v), f(x_i, y_v))) && (22)\\
&\quad \leqslant \Pr_{\{\boldsymbol{a}_{v,l}\}}(\forall v \in R, \forall l \in [L-1], \boldsymbol{a}_{v,l} \notin I) && (23)\\
&\quad = \Pr_{\{\boldsymbol{a}_{v',l}\}_{l\in[L-1]}}(\forall l \in [L-1], \boldsymbol{a}_{v',l} \notin I)^k && (24)\\
&\quad = [1 - \Pr_{\{\boldsymbol{a}_{v',l}\}_{l\in[L-1]}}(\exists l \in [L-1] \text{ s.t. } \boldsymbol{a}_{v',l} \in I)]^k \\
&\quad \leqslant (1 - \gamma(\frac{\epsilon}{4}))^k && (25)\\
&\quad \leqslant \exp(-\gamma(\frac{\epsilon}{4})k). && (26)
\end{aligned}$$

Line (22) follows from the independence of $\{\boldsymbol{y}_v\}_{v\in[n_2]}$, $\boldsymbol{\Omega}$, and $\{\boldsymbol{x}_s\}_{s\in[n_1]}$ from $\{\boldsymbol{a}_{v,l}\}_{v\in[n_2],l\in[L-1]}$. Line (23) follows from the definition of I in line (21) and because the monotonic functions $\{\boldsymbol{g}_v\}_{v\in[n_2]}$ are identically distributed. Line (24) follows since $\{\boldsymbol{g}_v\}_{v\in R}$ are i.i.d., line (25) follows from the definition of γ, and line (26) follows from the inequality $1 - x \leqslant \exp(-x)$. Note that since $\mathcal{P}_{\mathcal{G}}$ is diverse by hypothesis, by Lemma D.3, $\gamma(\frac{\epsilon}{4}) > 0$.

Since $\{\boldsymbol{y}_v\}_{v\in[n_2]}$, $\boldsymbol{\Omega} \cup \{\boldsymbol{x}_s\}_{s\in[n_1]}$, and $\{\boldsymbol{a}_{v,l}\}_{v\in[n_2],l\in[L-1]}$ are independent and almost every $y \in \mathcal{Y}$ is r-discerning and $(\frac{\epsilon}{4},\delta)$-discriminative, taking the expectation of line (26) with respect to $\{\boldsymbol{y}_v\}_{v\in[n_2]}$, $\boldsymbol{\Omega}$, and $\{\boldsymbol{x}_s\}_{s\in[n_1]}$ over $A \cap B \cap E$ finishes the proof.

□

Lemma D.12 gives a bound on the probability of $D^{\epsilon}_{u,i,j}$ conditional on $A \cap B \cap C \cap E \cap M$ (defined in the proof of Theorem C.1).

Lemma D.12. *Under the setting described in Theorem C.1, let $u \in [n_2]$ and $i \neq j \in [n_1]$. Then,*

$$\Pr(D^{\epsilon}_{u,i,j}|A,B,C,E,M) = 0.$$

Proof. Define the sets

$$E_1 = \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) + \epsilon < f(\boldsymbol{x}_j, \boldsymbol{y}_u)\}$$
$$E_2 = \{f(\boldsymbol{x}_i, \boldsymbol{y}_u) > f(\boldsymbol{x}_j, \boldsymbol{y}_u) + \epsilon\}.$$

Then, by the union bound and the law of total probability,

$$\begin{aligned}\Pr(D^{\epsilon}_{u,i,j}|A,B,C,E,M) &\leqslant \Pr(\mathrm{PR}(u,i,j,\beta,k) = 1 \cap E_1|A,B,C,E,M) \\ &+ \Pr(\mathrm{PR}(u,i,j,\beta,k) = 0 \cap E_2|A,B,C,E,M) \\ &\leqslant \Pr(\mathrm{PR}(u,i,j,\beta,k) = 1|A,B,C,E_1,M) \\ &+ \Pr(\mathrm{PR}(u,i,j,\beta,k) = 0|A,B,C,E_2,M).\end{aligned}$$

The argument for bounding each of these terms is similar, so we only bound $\Pr(\mathrm{PR}(u,i,j,\beta,k) = 1|A,B,C,E_1,M)$.

Fix $\{\boldsymbol{y}_v = y_v\}_{v\in[n_2]}$ r-discerning and $(\frac{\epsilon}{4},\delta)$-discriminative, $\{\boldsymbol{x}_s = x_s\}_{s\in[n_1]}$, $\boldsymbol{\Omega} = \Omega$, and $\{\boldsymbol{a}_{v,l} = a_{v,l}\}_{v\in[n_2],l\in[L-1]}$ such that $A\cap B\cap C\cap E_1\cap M$ occurs . We claim that the set V in Pairwise-Rank consists of $v_1,\ldots,v_k \in W^{i,j}_u(\beta)$ such that for all $l \in [k]$, $y_{v_l} \in B_{\frac{\epsilon}{4}}(y_u)$. The event B implies that there are $v_1,\ldots,v_k$ such that for all $l \in [k]$, $\rho'(y_u, y_{v_l}) \geqslant 1 - \frac{\delta}{2}$. Then, since y_u is $(\frac{\epsilon}{4},\delta)$-discriminative, it follows that $y_{v_1},\ldots,y_{v_k} \in B_{\frac{\epsilon}{4}}(y_u)$. Suppose that $w \in W^{i,j}_u(\beta)$ such that $y_w \in B_{\frac{\epsilon}{4}}(y_u)^c$. Then, since y_u is $(\frac{\epsilon}{4},\delta)$-discriminative, it follows that that $\rho'(y_u,y_w) < 1-\delta$. Then, for all $l \in [k]$,

$$\begin{aligned}R_{w,u} &\leqslant \rho'(y_w,y_u) + \frac{\delta}{4} &(27)\\ &< 1 - \frac{3}{4}\delta \\ &\leqslant \rho'(y_u,y_{v_l}) - \frac{\delta}{4} \\ &\leqslant R_{u,v_l} &(28)\end{aligned}$$

where lines (27) and (28) follow by event C and $v_l, w \in W^{i,j}_u(\beta)$. Thus, Pairwise-Rank selects $v_1,\ldots,v_k \in W^{i,j}_u(\beta)$ such that for all $l\in[k]$, $y_{v_l} \in B_{\frac{\epsilon}{4}}(y_u)$. Thus, the claim follows.

Event E_1 implies that $f(x_i,y_u)+\epsilon < f(x_j,y_u)$. Fix $l \in [k]$. Then, by the Lipschitzness of f,

$$f(x_i,y_{v_l}) \leqslant f(x_i,y_u) + \frac{\epsilon}{4} < f(x_j,y_u) - \frac{3\epsilon}{4} \leqslant f(x_j,y_{v_l}) - \frac{\epsilon}{2}.$$

Hence, $\forall l \in [k]$, $f(x_i,y_{v_l}) + \frac{\epsilon}{2} < f(x_j,y_{v_l})$ and $h_{v_l}(x_i,y_{v_l}) \leqslant h_{v_l}(x_j,y_{v_l})$. Then, event M implies that there is some $l\in[k]$ such that $h_{v_l}(x_i,y_{v_l}) < h_{v_l}(x_j,y_{v_l})$. Thus, the majority vote outputs the correct result. Thus,

$$\begin{aligned}\Pr(\mathrm{PR}(u,i,j,\beta,k) = 1|A,B,C,E_1,M,&\\ \{\boldsymbol{y}_v = y_v\}_{v\in[n_2]}, \{\boldsymbol{x}_s = x_s\}_{s\in[n_1]},&\\ \boldsymbol{\Omega} = \Omega, \{\boldsymbol{a}_{v,l} = a_{v,l}\}_{v\in[n_2],l\in[L-1]}) = 0.& \qquad (29)\end{aligned}$$

Since line (29) holds for all $\{y_v\}_{v\in[n_2]}$ r-discerning and $(\frac{\epsilon}{4},\delta)$-discriminative, $\{a_{v,l}\}_{v\in[n_2],l\in[L-1]}$, $\{x_s\}_{s\in[n_1]}$, Ω conditioned on the set the set $A\cap B\cap C\cap E_1\cap M$ and almost every $y \in \mathcal{Y}$ is r-discerning and $(\frac{\epsilon}{4},\delta)$-discriminative, the result follows. □

E Proofs for Section 6

Proof of Theorem 3. By compactness of $\mathcal{Y}$, there exists a finite subcover $\{C_1, \ldots, C_n\}$ of $\mathcal{Y}$ where each open ball C_i has diameter $\frac{\epsilon}{2}$. Since by assumption, for all $r > 0$ and $y \in \mathcal{Y}$, $\mathcal{P}_{\mathcal{Y}}(B_r(y)) > 0$, we have that $\mathcal{P}_{\mathcal{Y}}(C_i) > 0$ for all $i = 1, \ldots, n$. Let Q_{n_2} denote the event that for every $l \in [n]$ and $i, j \in [n_1]$, there exists $u \in [n_2]$ such that $\boldsymbol{y}_u \in C_l$ and we observe $(i, u) \in \Omega$ and $(j, u) \in \boldsymbol{\Omega}$. Since $p > 0$, as $n_2 \longrightarrow \infty$, $\Pr(Q_{n_2}) \longrightarrow 1$.

Let $\{\boldsymbol{x}_i = x_i\}_{i \in [n_1]}$, $\{\boldsymbol{y}_u = y_u\}_{u \in [n_2]}$, and $\boldsymbol{\Omega} = \Omega$ such that Q_{n_2} occurs. Let $\sigma \in \mathcal{S}^{n_1 \times n_2}$ be an $\frac{\epsilon}{2}$-consistent minimizer of $\widehat{\text{dis}}(\cdot, H)$ over the sample. Towards a contradiction, suppose there exists y_u and $i \neq j \in [n_1]$ such that $\sigma(i, u) < \sigma(j, u)$, $h_u(x_i, y_u) > h_u(x_j, y_u)$, and $f(x_i, y_u) > f(x_j, y_v) + \epsilon$. Without loss of generality, suppose that $y_u \in C_1$.

Since Q_{n_2} occurs by assumption, there exists $v \in [n_2]$ such that $y_v \in C_1$ and $(i, v), (j, v) \in \Omega$. Since σ is an $\frac{\epsilon}{2}$-consistent collection of rankings and the diameter of C_1 is $\frac{\epsilon}{2}$, σ gives the same ranking to y_u and y_v. Then, since $\sigma(i, u) < \sigma(j, u)$, it follows that $\sigma(i, v) < \sigma(j, v)$. By Lipschitzness of f,

$$f(x_i, y_v) \geqslant f(x_i, y_u) - \frac{\epsilon}{2} > f(x_j, y_u) + \frac{\epsilon}{2} \geqslant f(x_j, y_v). \tag{30}$$

Since g_v is strictly increasing, line (30) implies that $h_v(x_i, y_v) > h_v(x_j, y_v)$. Thus, σ is not a minimizer of $\widehat{\text{dis}}(\cdot, H)$–a contradiction. Thus, $\forall u \in [n_2]$ and $i \neq j \in [n_1]$ if $\sigma(i, u) < \sigma(j, u)$ and $h_u(x_i, y_u) > h_u(x_j, y_u)$, then $f(x_i, y_u) \leqslant f(x_j, y_u) + \epsilon$, implying that $\text{dis}_\epsilon(\sigma, H) = 0$. ☐

Proof of Theorem 4. Fix $\{\boldsymbol{x}_i = x_i\}_{i \in [n_1]}$. By compactness of $\mathcal{Y}$, there exists a finite subcover $\{C_1, \ldots, C_n\}$ of $\mathcal{Y}$ where each open ball C_i has diameter $\frac{\epsilon}{8}$. For every $l \in [n]$, fix $z_l \in C_l$ and define $P_l = \{(i, j) : f(x_i, z_l) > f(x_j, z_l) + \frac{\epsilon}{2}\}$.

Fix $l \in [n]$ and $(i, j) \in P_l$. Let $Q_{n_2}^{l,i,j}$ denote the event that there exists $\boldsymbol{y}_u \in C_l$ with $(i, u), (j, u) \in \boldsymbol{\Omega}$ and $\boldsymbol{a}_{u,q} \in (f(x_j, \boldsymbol{y}_u), f(x_i, \boldsymbol{y}_u))$ for some $q \in [L-1]$. Further, define

$$Q_{n_2} = \cap_{l \in [n], (i,j) \in P_l} Q_{n_2}^{l,i,j}.$$

Observe that by the Lipschitzness of f, for every $z \in C_l$, if $(i, j) \in P_l$, then $f(x_i, z) > f(x_j, z) + \frac{\epsilon}{4}$. Since n is fixed and finite, $|P_l|$ is fixed and finite, and the probability of observing a rating, p, is fixed, there exists a positive constant $C > 0$ such that $\Pr_{\boldsymbol{y}_u, \boldsymbol{\Omega}}(Q_{n_2}^{l,i,j} \mid \{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}) \geqslant C$. Thus, $\Pr(Q_{n_2}^{l,i,j} \mid \{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}) \longrightarrow 1$ as $n_2 \longrightarrow \infty$. Then, by the union bound,

$$\begin{aligned}\lim_{n_2 \longrightarrow \infty} \Pr_{\boldsymbol{y}_u, \boldsymbol{\Omega}}([Q_{n_2}]^c \mid \{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}) &\leqslant \lim_{n_2 \longrightarrow \infty} n \binom{n_1}{2} \Pr_{\boldsymbol{y}_u, \boldsymbol{\Omega}}([Q_{n_2}^{l,i,j}]^c \mid \{\boldsymbol{x}_s = x_s\}_{s \in [n_1]}) \\ &= 0.\end{aligned}$$

Since $\mathbb{E}[\mathbf{1}\{Q_{n_2}\} | \{\boldsymbol{x}_i\}_{i \in [n_1]}] \leqslant 1$, by the dominated convergence theorem,

$$\begin{aligned}\lim_{n_2 \longrightarrow \infty} \Pr(Q_{n_2}) &= \lim_{n_2 \longrightarrow \infty} \mathbb{E}_{\{\boldsymbol{x}_i\}} \mathbb{E}[\mathbf{1}\{Q_{n_2}\} | \{\boldsymbol{x}_i\}_{i \in [n_1]}] \\ &= \mathbb{E}_{\{\boldsymbol{x}_i\}} \lim_{n_2 \longrightarrow \infty} \mathbb{E}[\mathbf{1}\{Q_{n_2}\} | \{\boldsymbol{x}_i\}_{i \in [n_1]}] \\ &= 1\end{aligned}$$

Now, condition on $\{\boldsymbol{x}_i = x_i\}_{i\in[n_1]}, \{\boldsymbol{y}_u = y_u\}_{u\in[n_2]}, \boldsymbol{\Omega} = \Omega, \{\boldsymbol{a}_{u,l} = a_{u,l}\}_{u\in[n_2],l\in[L-1]}$ such that Q_{n_2} happens. Let $\sigma \in \mathcal{S}^{n_1\times n_2}$ be an $\frac{\epsilon}{8}$-consistent minimizer of $\widehat{\mathrm{dis}}(\cdot, H)$. Towards a contradiction, suppose there exists y_u and $i \neq j \in [n_1]$ such that $\sigma(i,u) < \sigma(j,u)$, $h_u(x_i, y_u) > h_u(x_j, y_u)$, and $f(x_i, y_u) > f(x_j, y_v) + \epsilon$. Without loss of generality, suppose that $y_u \in C_1$. We have that $(i,j) \in P_1$ since

$$
\begin{aligned}
f(x_i, z_1) &\geqslant f(x_i, y_u) - \frac{\epsilon}{8} \\
&\geqslant f(x_j, y_u) + \frac{7}{8}\epsilon \\
&\geqslant f(x_j, z_1) + \frac{3}{4}\epsilon.
\end{aligned}
$$

Therefore, the event Q_{n_2} implies that there exists $y_v \in C_1$ such that $(i,v),(j,v) \in \Omega$ and there exists $a_{v,q} \in (f(x_j, y_v), f(x_i, y_v))$. By the Lipschitzness of f, $f(x_j, y_v) < f(x_i, y_v)$, so that $h(x_j, y_v) < h(x_i, y_v)$. Since σ is $\frac{\epsilon}{8}$-consistent, $\sigma(i,v) < \sigma(j,v)$. But, then σ is not a minimizer of $\widehat{\mathrm{dis}}(\cdot, H)$ over the sample–a contradiction. Thus, $\forall u \in [n_2]$ and $i \neq j \in [n_1]$ if $\sigma(i,u) < \sigma(j,u)$ and $h_u(x_i, y_u) > h_u(x_j, y_u)$, then $f(x_i, y_u) \leqslant f(x_j, y_u) + \epsilon$, implying that $\mathrm{dis}_\epsilon(\sigma, H) = 0$. □

Proof of Theorem 5. Let $\boldsymbol{x}_1 = x_1, \ldots, \boldsymbol{x}_{n_1} = x_{n_1}, \boldsymbol{y}_1 = y_1, \ldots, \boldsymbol{y}_{n_2} = y_{n_2}$. Towards a contradiction, suppose that σ is not an ϵ-consistent collection of rankings over T. Then, there exists $i,j \in [n_1]$ and $u, v \in [n_2]$ such that $(i,j,u),(i,j,v) \in T$ and

$$
\begin{aligned}
&d_{\mathcal{Y}}(y_u, y_v) \leqslant \epsilon, &(31)\\
&\sigma(j,u) < \sigma(i,u), &(32)\\
&\sigma(j,v) > \sigma(i,v). &(33)
\end{aligned}
$$

Further, by definition of T,

$$
\begin{aligned}
&|f(x_j, y_u) - f(x_i, y_u)| > \epsilon &(34)\\
&|f(x_i, y_v) - f(x_j, y_v)| > \epsilon. &(35)\\
&h(x_i, y_u) \neq h(x_j, y_u) &(36)\\
&h(x_i, y_v) \neq h(x_j, y_v) &(37)
\end{aligned}
$$

Since $\mathrm{dis}_\epsilon(\sigma, H) = 0$ by hypothesis, and by inequalities (32), (33), (34), (35), (36), and (37) it follows that $h(x_j, y_u) < h(x_i, y_u)$ and $h(x_i, y_v) < h(x_j, y_v)$. Thus, by monotonicity of g_u, g_v,

$$
\begin{aligned}
\epsilon + f(x_j, y_u) &< f(x_i, y_u),\\
\epsilon + f(x_i, y_v) &< f(x_j, y_v).
\end{aligned}
$$

Then,

$$
\begin{aligned}
f(x_i, y_u) - f(x_i, y_v) &= f(x_i, y_u) - f(x_j, y_u) + f(x_j, y_u) - f(x_j, y_v) + f(x_j, y_v) - f(x_i, y_v)\\
&> 2\epsilon + f(x_j, y_u) - f(x_j, y_v).
\end{aligned}
$$

Then, rearranging the above equation and applying the Lipschitzness of f, we have that

$$2\epsilon < f(x_j, y_v) - f(x_j, y_u) + f(x_i, y_u) - f(x_i, y_v) \leqslant 2d_{\mathcal{Y}}(y_v, y_u),$$

which contradicts inequality (31). □

F Proof of Proposition 2 and other Results

In the following proposition, we give a simple illustrative example of a 1-Lipschitz function that is (ϵ, δ)-discriminative and r-discerning.

Proposition F.1. *Let $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]$, $\mathcal{P}_\mathcal{X}$ be the Lebesgue measure over $\mathcal{X}$, and $\mathcal{P}_\mathcal{Y}$ be the Lebesgue measure over $\mathcal{Y}$. Suppose that for all $u \in [n_2]$, g_u is strictly increasing. Consider the function*

$$f(x,y) = \begin{cases} x & : x \in [0,y] \\ y - x & : x \in (y,1] \end{cases}$$

Then, for all $1 > \epsilon > 0$, every $y \in \mathcal{Y}$ is (ϵ, ϵ^2)-discriminative. Further, there exists a positive nondecreasing r such that $\lim_{r \longrightarrow 0} r(z) = 0$ and every $y \in \mathcal{Y}$ is r-discerning.

Proof. Let $\epsilon \in (0,1)$ and suppose that $|y_1 - y_2| = \epsilon$. Without loss of generality, suppose that $y_1 < y_2$. Then, when $x_1 < x_2 \in (y_1, y_1 + \epsilon)$, $f(x_1, y_1) > f(x_2, y_1)$ and $f(x_1, y_2) < f(x_2, y_2)$. Since g_u is strictly increasing, $h_1(x_1, y_1) > h_1(x_2, y_1)$ and $h_2(x_1, y_2) < h_2(x_2, y_2)$. Since $\mathcal{P}_\mathcal{X} \times \mathcal{P}_\mathcal{X}((y_1, y_1 + \epsilon) \times (y_1, y_1 + \epsilon)) = \epsilon^2$, it follows that $\rho(y_1, y_2) < 1 - \epsilon^2$.

Clearly, there exists a positive nondecreasing r such that $\lim_{r \longrightarrow 0} r(z) = 0$ and every $y \in \mathcal{Y}$ is r-discerning. □

This example can easily be generalized to $f(x,y) = \|x - y\|_2$. The following proposition shows that by adding a dimension, the model $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^t \boldsymbol{y}$ with $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is a special case of the model $f(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = \|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{y}}\|_2$ with $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \mathbb{R}^{d+1}$. A similar construction in the other direction exists.

Proposition F.2. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1} \in \mathbb{R}^d$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2} \in \mathbb{R}^d$. There exist $\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_{n_1} \in \mathbb{R}^{d+1}$ and $\tilde{\boldsymbol{y}}_1, \ldots, \tilde{\boldsymbol{y}}_{n_2} \in \mathbb{R}^{d+1}$ such that $\forall u \in [n_2]$ and $\forall i \neq j \in [n_1]$, $\boldsymbol{x}_i^t \boldsymbol{y}_u > \boldsymbol{x}_j^t \boldsymbol{y}_u$ if and only if $\|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}_u\|_2 > \|\tilde{\boldsymbol{x}}_j - \tilde{\boldsymbol{y}}_u\|_2$.*

Proof. Let $B = \max_{i \in [n_1]} \|\boldsymbol{x}_i\|_2$. For all $i \in [n_1]$, there exists $\gamma_i \geqslant 0$ such that $\tilde{\boldsymbol{x}}_i := (\boldsymbol{x}_i^t, \gamma_i)^t$ and $\|\tilde{\boldsymbol{x}}_i\|_2 = B$ (by continuity and monotonicity of $\|\cdot\|_2$). For all $u \in [n_2]$, define $\tilde{\boldsymbol{y}}_u = (-\boldsymbol{y}_u^t, 0)^t$.

Fix $u \in [n_2]$ and $i \neq j \in [n_1]$. Then,

$$\begin{aligned} \|\tilde{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}_u\|_2^2 - \|\tilde{\boldsymbol{x}}_j - \tilde{\boldsymbol{y}}_u\|_2^2 &= \|\tilde{\boldsymbol{x}}_i\|_2^2 + \|\tilde{\boldsymbol{y}}_u\|_2^2 - 2\tilde{\boldsymbol{x}}_i^t \tilde{\boldsymbol{y}}_u - (\|\tilde{\boldsymbol{x}}_j\|_2^2 + \|\tilde{\boldsymbol{y}}_u\|_2^2 - 2\tilde{\boldsymbol{x}}_j^t \tilde{\boldsymbol{y}}_u) \\ &= -2\tilde{\boldsymbol{x}}_i^t \tilde{\boldsymbol{y}}_u + 2\tilde{\boldsymbol{x}}_j^t \tilde{\boldsymbol{y}}_u \\ &= \boldsymbol{x}_i^t \boldsymbol{y}_u - \boldsymbol{x}_j^t \boldsymbol{y}_u. \end{aligned}$$

The result follows. □

Proof of Proposition 2. 1. Consider a fixed $y \in \mathcal{Y}$. Fix $\boldsymbol{x}_2 = x_2 \in \mathcal{X}$. Then,

$$\begin{aligned} \Pr_{\boldsymbol{x}_1}(|\,\|\boldsymbol{x}_1 - y\|_2 - \|\boldsymbol{x}_2 - y\|_2\,| \leqslant 2\epsilon) &\leqslant \Pr_{\boldsymbol{x}_1}(\boldsymbol{x}_1 \in B_{\|\boldsymbol{x}_2 - y\| + 2\epsilon}(y) \backslash B_{\|\boldsymbol{x}_2 - y\| - 2\epsilon}(y)) \\ &\leqslant \sup_{z \in [0,2]} \mathcal{P}_\mathcal{X}(B_z(y) \backslash B_{z - 4\epsilon}(y)) \\ &= r(\epsilon) \end{aligned}$$

Taking the expectation with respect to $\boldsymbol{x}_2$ establishes the first part of this result.

Fix $y_u \in \mathcal{Y}$ and $\epsilon > 0$ and set $\delta = 2\mathcal{P}_{\mathcal{X}}(B_{\frac{\epsilon}{2}}(y_u))^2$. Fix $y_v \in B_\epsilon(y_u)^c \cap \mathcal{Y}$. If $\boldsymbol{x}_1 = x_1 \in B_{\frac{\epsilon}{2}}(y_u)$ and $\boldsymbol{x}_2 = x_2 \in B_{\frac{\epsilon}{2}}(y_v)$, then

$$[f(x_1, y_u) - f(x_2, y_u)][f(x_1, y_v) - f(x_2, y_v)] < 0.$$

A similar argument applies to the case $\boldsymbol{x}_1 = x_1 \in B_{\frac{\epsilon}{2}}(y_v)$ and $\boldsymbol{x}_2 = x_2 \in B_{\frac{\epsilon}{2}}(y_u)$. Thus, since by hypothesis, g_u is strictly increasing for all $u \in [n_2]$,

$$\rho(y_u, y_v) < 1 - \delta.$$

2. Both results follow immediately.

□

G Useful Bounds

Proposition G.1 (Chernoff-Hoeffding's Bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [a_i, b_i]$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Then,*

$$\Pr(|\bar{X} - \mathbb{E}[\bar{X}]| \geqslant t) \leqslant 2\exp(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}).$$

Proposition G.2 (Chernoff's multiplicative bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables with values in $[0, 1]$. Let $X = \sum_{i=1}^n X_i$. Then, for any $\epsilon > 0$,*

$$\Pr(X > (1+\epsilon)\mathbb{E}[X]) < \exp(-\frac{\epsilon^2\mathbb{E}[X]}{3}),$$
$$\Pr(X < (1-\epsilon)\mathbb{E}[X]) < \exp(-\frac{\epsilon^2\mathbb{E}[X]}{2}).$$

References

C. E. Lee, Y. Li, D. Shah, and D. Song. Blind regression: Nonparametric regression for latent variable models via collaborative filtering. *Advances in Neural Information Processing Systems*, 2016.