Supplementary materials: Magnitude-preserving ranking for structured outputs

Céline Brouard[1] Eric Bach[1] Sebastian Böcker[2] Juho Rousu[1]

[1] Helsinki Institute for Information Technology, Department of Computer Science, Aalto University, Espoo, Finland
[2] Chair for Bioinformatics, Friedrich-Schiller University, Jena, Germany

1 Proof of Theorem 1

Theorem 1. *The solution of the optimization problem:*

$$\underset{h\in\mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{\ell} \frac{1}{n_i} \sum_{j\in C_i} \|(h(x_i)-h(x_j))-(\psi(y_i)-\psi(y_j))\|^2_{\mathcal{F}_y} + \lambda\|h\|^2_{\mathcal{H}},\ \lambda>0, \tag{1}$$

admits a representation of the form:

$$\forall x\in\mathcal{X},\, h(x)=\sum_{i\in S\cup C}\mathcal{K}_x(x,x_i)\mathbf{c}_i,\, \mathbf{c}_i\in\mathcal{F}_y.$$

Proof. We consider the space of functions

$$\mathcal{H}_0=\{h\in\mathcal{H}|h(\cdot)=\sum_{i\in S\cup C}\mathcal{K}_x(\cdot,x_i)\mathbf{c}_i\}$$

and its orthogonal complement

$$\mathcal{H}_0^{\perp}=\{g\in\mathcal{H}|\langle g(\cdot),h(\cdot)\rangle_{\mathcal{H}}=0,\,\forall h\in\mathcal{H}_0\}.$$

Any function $h\in\mathcal{H}$ can be decomposed as a sum of two functions, one belonging to $\mathcal{H}_0$ and the other to $\mathcal{H}_0^{\perp}$: $h=h_0+h_0^{\perp}$.

By using the reproducing property we observe that the evaluation of a function h on any point $x\in S\cup C$ is independent of $h_0^{\perp}$:

$$\forall x\in S\cup C,\, h(x)=\langle h_0,\mathcal{K}_x(\cdot,x)\rangle_{\mathcal{H}}+\langle h_0^{\perp},\mathcal{K}_x(\cdot,x)\rangle_{\mathcal{H}}=h_0(x).$$

For the regularization term, we use the fact that $h_0^{\perp}$ is orthogonal to h_0 to show that:

$$\|h\|^2_{\mathcal{H}}=\|h_0\|^2_{\mathcal{H}}+\|h_0^{\perp}\|^2_{\mathcal{H}}.$$

Using these two properties, we can express the objective function as: $\mathcal{J}(h)=\mathcal{J}(h_0)+\lambda\|h_0^{\perp}\|^2_{\mathcal{H}}$. This means that the minimizer h of Equation (1) must have $h_0^{\perp}=0$ and therefore admits a representation of the form: $h(\cdot)=\sum_{i\in S\cup C}\mathcal{K}_x(\cdot,x_i)\mathbf{c}_i$, $\mathbf{c}_i\in\mathcal{F}_y$. □

2 Proof of Theorem 2

Theorem 2. *The optimization problem (5) can be rewritten under the following form:*

$$\mathcal{J}(h) = \lambda\|h\|_{\mathcal{H}}^2 + \sum_{j\in S\cup C} \|W\phi'(x_j) - \psi'(y_j)\|_{\mathcal{F}_y}^2. \tag{2}$$

The modified input feature vectors are defined as:

$$\phi'(x_j) = \begin{cases} \phi(x_j) - \overline{\phi}_{C_j} & \text{if } j \in S \\ \frac{1}{\sqrt{n_i}}\left(\phi(x_j) - \overline{\phi}_{C_i}\right) & \text{if } j \in C_i \end{cases},$$

where $\overline{\phi}_{C_i} = \frac{1}{n_i}\sum_{j\in C_i}\phi(x_j)$. *The modified output feature vectors are defined similarly.*

Proof. We show the equality between $\sum_{j\in\{i\}\cup C_i}\|W\phi'(x_j) - \psi'(y_j)\|_{\mathcal{F}_y}^2$ and $\frac{1}{n_i}\sum_{j\in C_i}\|(h(x_i) - h(x_j)) - (\psi(y_i) - \psi(y_j))\|_{\mathcal{F}_y}^2$:

$$\begin{aligned}
&\sum_{j\in\{i\}\cup C_i}\|W\phi'(x_j) - \psi'(y_j)\|_{\mathcal{F}_y}^2\\
=&\|W(\phi(x_i) - \overline{\phi}_{C_i}) - (\psi(y_i) - \overline{\psi}_{C_i})\|_{\mathcal{F}_y}^2 + \frac{1}{n_i}\sum_{j\in C_i}\|W(\phi(x_j) - \overline{\phi}_{C_i}) - (\psi(y_j) - \overline{\psi}_{C_i})\|_{\mathcal{F}_y}^2\\
=&\|(h(x_i) - \overline{h}_{C_i}) - (\psi(y_i) - \overline{\psi}_{C_i})\|_{\mathcal{F}_y}^2 + \frac{1}{n_i}\sum_{j\in C_i}\|(h(x_j) - \overline{h}_{C_i}) - (\psi(y_j) - \overline{\psi}_{C_i})\|_{\mathcal{F}_y}^2\\
=&\|h(x_i) - \psi(y_i)\|_{\mathcal{F}_y}^2 + \|\overline{h}_{C_i} - \overline{\psi}_{C_i}\|_{\mathcal{F}_y}^2 - 2(h(x_i) - \psi(y_i))^T(\overline{h}_{C_i} - \overline{\psi}_{C_i})\\
&+\frac{1}{n_i}\sum_{j\in C_i}\Big(\|h(x_j) - \psi(y_j)\|_{\mathcal{F}_y}^2 + \|\overline{h}_{C_i} - \overline{\psi}_{C_i}\|_{\mathcal{F}_y}^2 - 2(h(x_j) - \psi(y_j))^T(\overline{h}_{C_i} - \overline{\psi}_{C_i})\Big)\\
=&\|h(x_i) - \psi(y_i)\|_{\mathcal{F}_y}^2 + \|\overline{h}_{C_i} - \overline{\psi}_{C_i}\|_{\mathcal{F}_y}^2 - 2(h(x_i) - \psi(y_i))^T(\overline{h}_{C_i} - \overline{\psi}_{C_i})\\
&+\frac{1}{n_i}\sum_{j\in C_i}\|h(x_j) - \psi(y_j)\|_{\mathcal{F}_y}^2 + \|\overline{h}_{C_i} - \overline{\psi}_{C_i}\|_{\mathcal{F}_y}^2 - 2(\sum_{j\in C_i}h(x_j) - \psi(y_j))^T(\overline{h}_{C_i} - \overline{\psi}_{C_i})\\
=&\|h(x_i) - \psi(y_i)\|_{\mathcal{F}_y}^2 + \frac{1}{n_i}\sum_{j\in C_i}\|h(x_j) - \psi(y_j)\|_{\mathcal{F}_y}^2 - 2(h(x_i) - \psi(y_i))^T(\overline{h}_{C_i} - \overline{\psi}_{C_i})\\
=&\frac{1}{n_i}\sum_{j\in C_i}\|(h(x_i) - \psi(y_i)) - (h(x_j) - \psi(y_j))\|_{\mathcal{F}_y}^2\\
=&\frac{1}{n_i}\sum_{j\in C_i}\|(h(x_i) - h(x_j)) - (\psi(y_i) - \psi(y_j))\|_{\mathcal{F}_y}^2.
\end{aligned}$$

□

3 Proof of Proposition 4

When replacing the matrices containing candidate input feature vectors by their approximation given in Equation (9): $\Phi_C = \Phi_S M \Psi_S^T \Psi_C$ and $\bar{\Phi}_C = \Phi_S M \Psi_S^T \bar{\Psi}_C$, the expression of Φ' becomes:

$$\begin{aligned}
\Phi' &= \left[\Phi_S - \Phi_S M\Psi_S^T\bar{\Psi}_C, (\Phi_S M\Psi_S^T\Psi_C - \Phi_S M\Psi_S^T\bar{\Psi}_C V^T)D_n\right]\\
&= \Phi_S\left[I_\ell - M\Psi_S^T\bar{\Psi}_C, M\Psi_S^T(\Psi_C - \bar{\Psi}_C V^T)D_n\right].
\end{aligned}$$

We note $A = \left[I_\ell - M\Psi_S^T\bar{\Psi}_C, M\Psi_S^T(\Psi_C - \bar{\Psi}_C V^T)D_n\right]$ and replace the expression of Φ' in the solution:

$$W = \Psi'\left(\lambda I_{\ell+n} + A^T K_{X_S} A\right)^{-1} A^T \Phi_S^T = \Psi' A^T\left(\lambda I_\ell + K_{X_S} A A^T\right)^{-1}\Phi_S^T.$$

4 Proof of Proposition 5

We replace the candidate input feature vectors by their approximation and use the fact that the training input feature vectors can be expressed as: $\Phi_S = D_{\Phi_S}\mathbf{I}$.

$$\begin{aligned}\Phi' &= \left[\Phi_S - \bar{\Phi}_C, (\Phi_C - \bar{\Phi}_C V^T)D_n\right]\\ &= \left[D_{\Phi_S}\mathbf{I} - D_{\Phi_S}\mathbf{M}\Psi_S^T\bar{\Psi}_C, (D_{\Phi_S}\mathbf{M}\Psi_S^T\Psi_C - D_{\Phi_S}\mathbf{M}\Psi_S^T\bar{\Psi}_C V^T)D_n\right]\\ \Phi' &= D_{\Phi_S}\left[\mathbf{I} - \mathbf{M}\Psi_S^T\bar{\Psi}_C, \mathbf{M}\Psi_S^T(\Psi_C - \bar{\Psi}_C V^T)D_n\right].\end{aligned}$$

We note $A = \left[\mathbf{I} - \mathbf{M}\Psi_S^T\bar{\Psi}_C, \mathbf{M}\Psi_S^T(\Psi_C - \bar{\Psi}_C V^T)D_n\right]$. Given the expression of Φ', the solution can be rewritten as follows:

$$W = \Psi'\left(\lambda I_{\ell+n} + A^T D_{\Phi_S}^T D_{\Phi_S} A\right)^{-1} A^T D_{\Phi_S}^T = \Psi' A^T \left(\lambda I_{\ell K} + D_{K_{X_S}} A A^T\right)^{-1} D_{\Phi_S}^T,$$

where $D_{K_{X_S}} = \operatorname{diag}(\mu_1 K_{X_S}^1, \ldots, \mu_k K_{X_S}^K)$.