Active Learning With Uniform Feature Noise : Appendix

1 Justifying Claims in the Lower Bounds

Approximations:

1. $(x+y)^k = x^k(1+y/x)^k \approx x^k + kx^{k-1}y$ when $y \prec x$. Even when $y \preceq x$, both terms are the same order.
2. $(x-y)^k = x^k(1-y/x)^k \approx x^k - kx^{k-1}y$ when $y \prec x$. Even when $y \preceq x$ both terms are the same order.
3. When $y < x$ but not $y \prec x$, by Taylor expansion of $(1+z)^k$ around $z=0$, we have $(x+y)^k = x^k(1+y/x)^k = x^k[1+(1+c)^{k-1}y/x] = x^k + Cx^{k-1}y$ for some $0 < c < y/x < 1$ and some constant C. Similarly for $(x-y)^k$.

Let's assume the boundary is at $-\sigma$ for easier calculations. (we denote a_n, σ_n as a, σ here). Remember

$$m_1(x) = 1/2 + cx|x|^{k-2} \text{ if } x \geq -\sigma$$

$$m_2(x) = \begin{cases} 1/2 + c(x-a)|x-a|^{k-2} & \text{if } x < \beta a + \sigma \\ m_1(x) & \text{if } x \geq \beta a + \sigma \end{cases}$$

where $\beta = \frac{1}{1-(c/C)^{1/(k-1)}} \geq 1$ is such that $m_2 \in P(\kappa, c, C, \sigma)$. Clearly, when $x < \beta a + \sigma$, m_2 satisfies condition (T). So, we only need to verify that whenever $x \geq \beta a + \sigma$ we have

$$m_2(x) - 1/2 \;=\; cx^{k-1} \;\leq\; C(x-a)^{k-1} \tag{1}$$

This statement holds iff $(c/C)^{1/(k-1)} \leq 1 - a/x \;\Leftrightarrow\; a/x \leq 1 - (c/C)^{1/(k-1)} \;\Leftrightarrow\; x \geq \beta a$, which holds for all $\sigma \geq 0$, and hence m_2 satisfies condition (T).

Proposition 1. When $\sigma \prec a$, $\max_w |F_1(w) - F_2(w)| \asymp a^{k-1}$
Proposition 2. When $\sigma \succ a$ $\max_w |F_1(w) - F_2(w)| \asymp \sigma^{k-2}a$

Let us now prove these two propositions, with detailed calculations in each case (note that when $\sigma \asymp a$, then $\max_w |F_1(w) - F_2(w)| \asymp a^{k-1} \asymp \sigma^{k-2}a$, and can be checked using our approximations 1,2,3).

1. When $\sigma \prec a$, we will prove proposition 1. Remember that we can't query in $-\sigma \leq w \leq 0$.

   (a) When $0 \leq w \leq \sigma$, we have

$$\begin{aligned} F_1(w) = (m_1 \star U)(w) &= \int_{w-\sigma}^{0} (1/2 - cx|x|^{k-2})dx/2\sigma + \int_0^{w+\sigma} (1/2 + cx^{k-1})dx/2\sigma && (2)\\ &= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (\sigma - w)^k] && (3)\\ &= 1/2 + \frac{c}{2\sigma k}\sigma^k[(1+w/\sigma)^k - (1-w/\sigma)^k] && (4)\\ &\approx 1/2 + c\sigma^{k-2}w && (5) \end{aligned}$$

$$\begin{aligned} F_2(w) = (m_2 \star U)(w) &= \int_{w-\sigma}^{w+\sigma} (1/2 - c(x-a)|x-a|^{k-2})dx/2\sigma && (6)\\ &= 1/2 - \frac{c}{2\sigma k}[(a-w-\sigma)^k - (a+\sigma-w)^k] && (7)\\ &\approx 1/2 - c(a-w)^{k-1} && (8)\\ & && (9) \end{aligned}$$

[Boundaries: $F_1(0) - \frac{1}{2} = 0, F_1(\sigma) - \frac{1}{2} \asymp \sigma^{k-1}, F_2(0) - \frac{1}{2} \asymp -a^{k-1}, F_2(\sigma) - \frac{1}{2} \asymp -a^{k-1}$].

$$F_1(w) - F_2(w) \;\preceq\; a^{k-1} \tag{10}$$

(b) When $\sigma \le w \le a - \sigma$

$$\begin{aligned} F_1(w) = (m_1 \star U)(w) &= \int_{w-\sigma}^{w+\sigma} (1/2 + cx^{k-1}) dx/2\sigma & (11) \\ &= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (w-\sigma)^k] & (12) \\ &\approx 1/2 + cw^{k-1} & (13) \end{aligned}$$

$$\begin{aligned} F_2(w) = (m_2 \star U)(w) &= \int_{w-\sigma}^{w+\sigma} (1/2 - c(x-a)|x-a|^{k-2}) dx/2\sigma & (14) \\ &= 1/2 - \frac{c}{2\sigma k}[(a-w-\sigma)^k - (a+\sigma-w)^k] & (15) \\ &\approx 1/2 - c(a-w)^{k-1} & (16) \end{aligned}$$

[Boundaries: $F_1(\sigma) - \frac{1}{2} \asymp \sigma^{k-1}, F_1(a-\sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(\sigma) - \frac{1}{2} \asymp -a^{k-1}, F_2(a-\sigma) - \frac{1}{2} \asymp -\sigma^{k-1}$].

$$\begin{aligned} F_1(w) - F_2(w) &= cw^{k-1} + c(a-w)^{k-1} & (17) \\ &\le c(a-\sigma)^{k-1} + c(a-\sigma)^{k-1} & (18) \\ &\preceq a^{k-1} & (19) \end{aligned}$$

(c) When $a - \sigma \le w \le a$

$$F_1(w) \approx 1/2 + cw^{k-1} \quad (20)$$

$$\begin{aligned} F_2(w) &= \int_{w-\sigma}^{a} (1/2 - c(x-a)|x-a|^{k-2}) dx/2\sigma + \int_{a}^{w+\sigma} 1/2 + c(x-a)^{k-1} dx/2\sigma & (21) \\ &= 1/2 - \frac{c}{2\sigma k}[(a-w+\sigma)^k - (w+\sigma-a)^k] & (22) \\ &\approx 1/2 - c\sigma^{k-2}(a-w) & (23) \end{aligned}$$

[Boundaries: $F_1(a-\sigma) - \frac{1}{2} \asymp a^{k-1}, F_1(a) - \frac{1}{2} \asymp a^{k-1}, F_2(a-\sigma) - \frac{1}{2} \asymp -\sigma^{k-1}, F_2(a) - \frac{1}{2} = 0$]

$$\begin{aligned} F_1(w) - F_2(w) &\approx cw^{k-1} + c\sigma^{k-2}(a-w) & (24) \\ &\le ca^{k-1} + c\sigma^{k-2}\sigma & (25) \\ &\preceq a^{k-1} & (26) \end{aligned}$$

(d) When $a \le w \le a + \sigma$

$$F_1(w) \approx 1/2 + cw^{k-1} \quad (27)$$

$$F_2(w) \approx 1/2 + c\sigma^{k-2}(a-w) \quad (28)$$

[Boundaries: $F_1(a) - \frac{1}{2} \asymp a^{k-1}, F_1(a+\sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(a) - \frac{1}{2} = 0, F_2(a+\sigma) - \frac{1}{2} \asymp \sigma^{k-1}$]

$$F_1(w) - F_2(w) \preceq a^{k-1}$$

(e) When $a + \sigma \le w \le \beta a - \sigma$

$$F_1(w) \approx 1/2 + cw^{k-1} \quad (29)$$

$$\begin{aligned} F_2(w) &= \int_{w-\sigma}^{w+\sigma} 1/2 + c(x-a)^{k-1} dx/2\sigma & (30) \\ &= 1/2 + \frac{c}{2\sigma k}[(w+\sigma-a)^k - (w-\sigma-a)^k] & (31) \\ &\approx 1/2 + c(w-a)^{k-1} & (32) \end{aligned}$$

[B: $F_1(a+\sigma) - \frac{1}{2} \asymp a^{k-1}, F_1(\beta a - \sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(a+\sigma) - \frac{1}{2} \asymp \sigma^{k-1}, F_2(\beta a - \sigma) - \frac{1}{2} \asymp a^{k-1}$]

$$\begin{aligned}
F_1(w) - F_2(w) &\approx cw^{k-1} - c(w-a)^{k-1} && (33)\\
&\leq c(\beta a - \sigma)^{k-1} + c\sigma^{k-1} && (34)\\
&\leq c(\beta^{k-1}+1)a^{k-1} && (35)\\
&\preceq a^{k-1} && (36)
\end{aligned}$$

(f) When $\beta a - \sigma \leq w \leq \beta a + \sigma$

$$F_1(w) \approx 1/2 + cw^{k-1} \qquad (37)$$

$$\begin{aligned}
F_2(w) &= \int_{w-\sigma}^{\beta a} 1/2 + c(x-a)^{k-1} dx/2\sigma + \int_{\beta a}^{w+\sigma} 1/2 + x^{k-1} dx/2\sigma && (38)\\
&= 1/2 + \frac{c}{2\sigma k}[(\beta a - a)^k - (w-\sigma-a)^k + (w+\sigma)^k - (\beta a)^k] && (39)
\end{aligned}$$

[$F_1(\beta a - \sigma) - \frac{1}{2} \asymp a^{k-1}, F_1(\beta a + \sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(\beta a - \sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(\beta a + \sigma) - \frac{1}{2} \asymp a^{k-1}$]

$$\begin{aligned}
F_1(w) - F_2(w) &= cw^{k-1} + \frac{c}{2\sigma k}[(\beta^k - (\beta-1)^k)a^k + (w-\sigma-a)^k - (w-\sigma)^k]\\
&\leq c(\beta+1)^{k-1}a^{k-1} + \frac{c}{2\sigma k}[(\beta a)^k - (\beta a - 2\sigma)^k] - \frac{c}{2\sigma k}[(\beta-1)^k a^k - ((\beta-1)a-\sigma)^k]\\
&\approx c(\beta+1)^{k-1}a^{k-1} + \frac{c}{2\sigma k}[k(\beta a)^{k-1}2\sigma] - \frac{c}{2\sigma k}[k(\beta-1)^{k-1}a^{k-1}\sigma]\\
&= ca^{k-1}[(\beta+1)^{k-1} + \beta^{k-1} - \tfrac{1}{2}(\beta-1)^{k-1}]\\
&\asymp a^{k-1}
\end{aligned}$$

(g) When $\beta a + \sigma \leq w \leq \beta a + 2\sigma$

$$F_1(w) = 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (w-\sigma)^k] \qquad (40)$$

$$\begin{aligned}
F_2(w) &= \int_{w-\sigma}^{\beta a+\sigma} 1/2 + c(x-a)^{k-1} dx/2\sigma + \int_{\beta a+\sigma}^{w+\sigma} 1/2 + cx^{k-1} dx/2\sigma\\
&= 1/2 + \frac{c}{2\sigma k}[(\beta a + \sigma - a)^k - (w-\sigma-a)^k + (w+\sigma)^k - (\beta a + \sigma)^k]
\end{aligned}$$

[$F_1(\beta a + \sigma) - \frac{1}{2} \asymp a^{k-1}, F_1(\beta a + 2\sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(\beta a + \sigma) - \frac{1}{2} \asymp a^{k-1}, F_2(\beta a + 2\sigma) - \frac{1}{2} \asymp a^{k-1}$]

$$\begin{aligned}
F_1(w) - F_2(w) &= \frac{c}{2\sigma k}[(\beta a + \sigma)^k - (\beta a + \sigma - a)^k + (w-\sigma-a)^k - (w-\sigma)^k] && (41)\\
&\approx \frac{c}{2\sigma k}[(\beta a + \sigma)^{k-1}ka - (w-\sigma)^{k-1}ka] && (42)\\
&\leq \frac{ca}{2\sigma}[(\beta a + \sigma)^{k-1} - (\beta a)^{k-1}] && (43)\\
&\approx \frac{ca}{2\sigma}[(\beta a)^{k-1}(1 + \frac{(k-1)\sigma}{\beta a}) - (\beta a)^{k-1}] && (44)\\
&= a^{k-1}[c\beta^{k-2}(k-1)/2] && (45)\\
&\asymp a^{k-1} && (46)
\end{aligned}$$

(h) When $w \geq \beta a + 2\sigma$

$$F_1(w) = F_2(w)$$

That completes the proof of the first claim.

2. When $\sigma \succ a$, we will prove the second proposition.

   (a) When $-\sigma \le w \le 0$, we are not allowed to query here.

   (b) When $0 < w \le \beta a$

$$
\begin{aligned}
F_1(w) = (m_1 \star U)(w) &= \int_{w-\sigma}^{0} (1/2 - cx|x|^{k-2})dx/2\sigma + \int_{0}^{w+\sigma} (1/2 + cx^{k-1})dx/2\sigma && (47)\\
&= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (\sigma - w)^k] && (48)\\
&= 1/2 + \frac{c}{2\sigma k}\sigma^k[(1+w/\sigma)^k - (1-w/\sigma)^k] && (49)\\
&\approx 1/2 + c\sigma^{k-2}w && (50)
\end{aligned}
$$

Similarly $F_2(w) \approx 1/2 + c\sigma^{k-2}(w-a)$

[Boundaries: $F_1(0) - \frac{1}{2} = 0, F_1(\beta a) - \frac{1}{2} \asymp \sigma^{k-2}a, F_2(0) - \frac{1}{2} \asymp -\sigma^{k-2}a, F_2(\beta a) \asymp \sigma^{k-2}a$]

$$F_1(w) - F_2(w) \asymp \sigma_n^{k-2}a.$$

   (c) When $\beta a \le w \le \sigma$

$$
\begin{aligned}
F_1(w) = &= \int_{w-\sigma}^{0} (1/2 - cx|x|^{k-2})dx/2\sigma + \int_{0}^{w+\sigma} (1/2 + cx^{k-1})dx/2\sigma && (51)\\
&= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (\sigma - w)^k] && (52)\\
&= 1/2 + \frac{c}{2\sigma k}\sigma^k[(1+w/\sigma)^k - (1-w/\sigma)^k] && (53)\\
&\approx 1/2 + c\sigma^{k-2}w && (54)
\end{aligned}
$$

$$
\begin{aligned}
F_2(w) &= \int_{w-\sigma}^{a} (1/2 - c(x-a)|x-a|^{k-2})\frac{dx}{2\sigma} + \int_{a}^{\beta a+\sigma} (1/2 + c(x-a)^{k-1})\frac{dx}{2\sigma} + \int_{\beta a+\sigma}^{w+\sigma} 1/2 + cx^{k-1}\frac{dx}{2\sigma}\\
&= 1/2 + \frac{c}{2\sigma k}[-(\sigma + a - w)^k + (\beta a + \sigma - a)^k + (w+\sigma)^k - (\beta a + \sigma)^k]\\
&\approx 1/2 + \frac{c}{2\sigma k}[-\sigma^k(1 - \frac{k(w-a)}{\sigma}) + \sigma^k(1 + \frac{k(\beta - 1)a}{\sigma}) + \sigma^k(1 + \frac{kw}{\sigma}) - \sigma^k(1 + \frac{k\beta a}{\sigma})]\\
&= 1/2 + \frac{c}{2}\sigma^{k-2}[w - a + (\beta - 1)a + w - \beta a]\\
&= 1/2 + c\sigma^{k-2}(w-a)
\end{aligned}
$$

[Boundaries: $F_1(\beta a) - \frac{1}{2} \asymp \sigma^{k-2}a, F_1(\sigma) - \frac{1}{2} \asymp \sigma^{k-1}, F_2(\beta a) \asymp \sigma^{k-2}a, F_2(\sigma) - \frac{1}{2} \asymp -\sigma^{k-2}a$]

$$F_1(w) - F_2(w) \asymp \sigma^{k-2}a$$

Specifically, verify the boundary at σ

$$
\begin{aligned}
F_1(\sigma) - F_2(\sigma) &= \frac{c}{2\sigma k}[a^k - (\beta a + \sigma - a)^k + (\beta a + \sigma)^k] && (55)\\
&= \frac{c}{2\sigma k}[a^k - \sigma^k(1 + k\frac{\beta a - a}{\sigma}) + \sigma^k(1 + k\frac{\beta a}{\sigma})] && (56)\\
&= \frac{c}{2\sigma k}[a^k + k\sigma^{k-1}a] && (57)\\
&\le c\sigma^{k-2}a && (58)
\end{aligned}
$$

(d) When $\sigma \leq w \leq a+\sigma$

$$F_1(w) = \int_{w-\sigma}^{w+\sigma} (1/2 + cx^{k-1})dx/2\sigma \quad (59)$$

$$= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (w-\sigma)^k] \quad (60)$$

$$(61)$$

$$F_2(w) = \int_{w-\sigma}^{a} (1/2 - c(x-a)|x-a|^{k-2})\frac{dx}{2\sigma} + \int_{a}^{\beta a+\sigma} (1/2 + c(x-a)^{k-1})\frac{dx}{2\sigma} + \int_{\beta a+\sigma}^{w+\sigma} 1/2 + cx^{k-1}\frac{dx}{2\sigma}$$
$$= 1/2 + \frac{c}{2\sigma k}[-(\sigma+a-w)^k + (\beta a+\sigma-a)^k + (w+\sigma)^k - (\beta a+\sigma)^k]$$

$$F_1(w) - F_2(w) = \frac{c}{2\sigma k}[(\sigma+a-w)^k - (\beta a+\sigma-a)^k - (w-\sigma)^k + (\beta a+\sigma)^k] \quad (62)$$

Differentiating the above term with respect to w, gives $\frac{c}{2\sigma}[-(\sigma+a-w)^{k-1} - (w-\sigma)^{k-1}] \leq 0$ because $\sigma \leq w \leq a+\sigma$ and hence $F_1(w) - F_2(w)$ is decreasing with w. We already saw $F_1(\sigma) - F_2(\sigma) \leq c\sigma^{k-2}a$. We can also verify that at the other boundary,

$$F_1(a+\sigma) - F_2(a+\sigma) = \frac{c}{2\sigma k}[-(\beta a+\sigma-a)^k - a^k + (\beta a+\sigma)^k] \quad (63)$$

$$= \frac{c}{2\sigma k}[-a^k - \sigma^k(1+k\frac{\beta a-a}{\sigma}) + \sigma^k(1+k\frac{\beta a}{\sigma})] \quad (64)$$

$$= \frac{c}{2\sigma k}[-a^k + k\sigma^{k-1}a] \quad (65)$$

$$\leq \frac{c}{2}\sigma^{k-2}a \quad (66)$$

(e) When $\sigma+a \leq w \leq \beta a+\sigma$

$$F_1(w) = \int_{w-\sigma}^{w+\sigma} (1/2 + cx^{k-1})dx/2\sigma \quad (67)$$

$$= 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (w-\sigma)^k] \quad (68)$$

$$(69)$$

$$F_2(w) = \int_{w-\sigma}^{\beta a+\sigma} (1/2 + c(x-a)^{k-1})\frac{dx}{2\sigma} + \int_{\beta a+\sigma}^{w+\sigma} 1/2 + cx^{k-1}\frac{dx}{2\sigma}$$
$$= 1/2 + \frac{c}{2\sigma k}[(\beta a+\sigma-a)^k - (w-\sigma-a)^k + (w+\sigma)^k - (\beta a+\sigma)^k]$$

$$F_1(w) - F_2(w) = \frac{c}{2\sigma k}[(w-\sigma-a)^k - (\beta a+\sigma-a)^k - (w-\sigma)^k + (\beta a+\sigma)^k] \quad (70)$$

$$(71)$$

Differentiating with respect to w gives $\frac{c}{2\sigma}[(w-\sigma-a)^{k-1} - (w-\sigma)^{k-1}] \leq 0$ because $w-\sigma-a \leq w-\sigma$ and so $F_1 - F_2$ is decreasing with w. We know $F_1(a+\sigma) - F_2(a+\sigma) \leq \frac{c}{2}\sigma^{k-2}a$, and we can verify at the other boundary that

$$F_1(\beta a+\sigma) - F_2(\beta a+\sigma) = \frac{c}{2\sigma k}[(\beta a-a)^k - (\beta a+\sigma-a)^k - (\beta a)^k + (\beta a+\sigma)^k] \quad (72)$$

$$\approx \frac{c}{2\sigma k}[(\beta a-a)^k - (\beta a)^k - \sigma^k(1+k\frac{\beta a-a}{\sigma}) + \sigma^k(1+k\frac{\beta a}{\sigma})] \quad (73)$$

$$= \frac{c}{2\sigma k}[(\beta a-a)^k - (\beta a)^k + k\sigma^{k-1}a] \quad (74)$$

$$\leq \frac{c}{2}\sigma^{k-2}a \quad (75)$$

(f) When $\beta a + \sigma \le w \le \beta a + 2\sigma$

$$F_1(w) = 1/2 + \frac{c}{2\sigma k}[(w+\sigma)^k - (w-\sigma)^k]$$

$$\begin{aligned} F_2(w) &= \int_{w-\sigma}^{\beta a+\sigma} 1/2 + c(x-a)^{k-1}dx/2\sigma + \int_{\beta a+\sigma}^{w+\sigma} 1/2 + cx^{k-1}dx/2\sigma && (76) \\ &= 1/2 + \frac{c}{2k\sigma}[(\beta a+\sigma-a)^k - (w-\sigma-a)^k + (w+\sigma)^k - (\beta a+\sigma)^k] && (77) \end{aligned}$$

Hence

$$\begin{aligned} F_1(w) - F_2(w) &= \frac{c}{2\sigma k}[(\beta a+\sigma)^k - (\beta a+\sigma-a)^k + (w-\sigma-a)^k - (w-\sigma)^k] && (78) \\ &\approx \frac{c}{2\sigma k}[(\beta a+\sigma)^{k-1}ka - (w-\sigma)^{k-1}ka] && (79) \\ &\le \frac{ca}{2\sigma}[(\beta a+\sigma)^{k-1} - (\beta a)^{k-1}] && (80) \\ &\approx c/2\sigma^{k-2}a && (81) \\ &\asymp \sigma^{k-2}a && (82) \end{aligned}$$

Alternately, by the same argument as in the previous case, differentiating with respect to w gives $\frac{c}{2\sigma}[(w-\sigma-a)^{k-1} - (w-\sigma)^{k-1}] \le 0$ because $w-\sigma-a \le w-\sigma$ and so $F_1 - F_2$ is decreasing with w. We know $F_1(\beta a+\sigma) - F_2(\beta a+\sigma) \le \frac{c}{2}\sigma^{k-2}a$, and we can verify at the other endpoint that

$$F_1(\beta a+2\sigma) - F_2(\beta a+2\sigma) = 0 \quad (83)$$

(g) When $w \ge \beta a + 2\sigma$, $F_1(w) = F_2(w)$

That completes the proof of the second proposition.

2 "Linear" Convolved Regression Function, Justifying Eq.(8,9,10,11)

For ease of presentation, let us assume the threshold is at 0, and define $m \in \mathcal{P}(c, C, k, \sigma)$ as

$$m(x) = \begin{cases} 1/2 + f(x) + \Delta(x) & \text{if } x \geq 0 \\ 1/2 - f(x) \text{ if } x < 0 \end{cases}$$

Due to assumption (M), $\Delta(x)$ must be 0 when $0 \leq x \leq \sigma$. Hence, the Taylor expansion of $\Delta(x)$ around $x = \sigma$ looks like

$$\Delta(x) = (x - \sigma)\Delta'(\sigma) + (x - \sigma)^2 \Delta''(\sigma) + \ldots$$

If one represents, as before, $F(x) = m \star U$, then directly from the definitions, it follows for $\delta > 0$ that

$$F(\delta) - F(0) = \int_{\sigma}^{\sigma+\delta} (1/2 + f(z) + \Delta(z)) \frac{dz}{2\sigma} - \int_{-\sigma}^{-\sigma+\delta} (1/2 - f(z)) \frac{dz}{2\sigma}$$

In particular, due to the form (T) of m, let $f = c_1 |x|^{k-1}$ for some $c \leq c_1 \leq C$ (we could also break f into parts where it has different c_1s but this is a technicality and does not change the behaviour). Then

$$\begin{aligned} F(\delta) - F(0) &= \frac{c_1}{2k\sigma}[(x^k)_{\sigma}^{\sigma+\delta} - (x^k)_{-\sigma}^{-\sigma+\delta}] + \int_{\sigma}^{\delta+\sigma} [(z-\sigma)\Delta'(\sigma) + (z-\sigma)^2\Delta''(\sigma) + \ldots] \frac{dz}{2\sigma} && (84) \\ &= \frac{c_1}{2k\sigma}[(\sigma+\delta)^k - \sigma^k + (-\sigma+\delta)^k - (-\sigma)^k] + \frac{[(z-\sigma)^2]_{\sigma}^{\sigma+\delta}}{4\sigma}\Delta'(\sigma) + \ldots && (85) \\ &\approx c_1 \sigma^{k-2}\delta + \frac{\delta^2}{4\sigma}\Delta'(\sigma) + o(\delta^2) && (86) \end{aligned}$$

Thus we get behaviour of the form

$$F(t + h) \geq 1/2 + c\sigma^{k-2} h$$

One can derive similar results when $\delta < 0$.

The claims about WIDEHIST immediately follow from the above, but we can make them a little more explicit. First note that $F(w) = 1/2 + \frac{c}{\sigma}(w - t)$ for w close to t (in fact for $w \in [t - \sigma, t + \sigma]$), as seen in Section 1 of this Appendix. Consider a bin just outside the bins $i^* - 1, i^*, i^* + 1$, for instance bin $i = i^* + 2$ centered at b_i (note $b_i \geq t + h$), and let J be the set of points j that fall within $b_i \pm \sigma/2$. Define

$$\widehat{p}_i = \frac{1}{n\sigma/2R} \sum_{j \in J} \mathbb{I}(Y_j = +)$$

where $Y_j \in \{\pm 1\}$ are observations at points $j \in J$. Now, we have, since $P(Y_j = +) = F(j)$

$$\begin{aligned} \mathbb{E}[\widehat{p}_i] &= \frac{1}{n\sigma/2R} \sum_{j \in J} F(j) \\ &= \frac{1}{n\sigma/2R} \left[\sum_{j \in J} 1/2 + \tfrac{c}{\sigma}(X_j - t) \right] \\ &\approx 1/2 + \frac{1}{\sigma} \int_{b_i - t - \sigma/2}^{b_i - t + \sigma/2} \tfrac{c}{\sigma} z \, dz \\ &= 1/2 + \frac{c}{2\sigma^2} \left[(b_i - t + \sigma/2)^2 - (b_i - t - \sigma/2)^2 \right] \\ &= 1/2 + \tfrac{c}{\sigma}(b_i - t) \\ &\geq 1/2 + \tfrac{c}{\sigma} h \end{aligned}$$

3 Justifying Claims in the Active Upper Bounds

Phase 1 ($k=1$)**.** In the first phase of the algorithm, it is possible that $\sigma \preceq R_e/n$ but $\succeq R_e e^{-n}$ - in other words the noise may be small enough that passive learning cannot make out that we are in the errors-in-variables setting, and then the passive estimator will get a point error of $\frac{C_1 R_e}{n/E}$ in each of those epochs (as if there is no feature noise). This point error is to the best point in epoch e, which we can prove by induction is the true threshold t with high probability. Since it trivially holds in the first epoch ($t \in D_1 = [-1,1]$), we assume that it is true in epoch $e-1$. Then, in epoch e, the true threshold t is still the best point if the estimator x_{e-1} of epoch $e-1$ was within R_e of t, or in other words if $|x_{e-1}-t| \leq R_e$. This would definitely hold if $\frac{C_1 R_{e-1}}{n/E} \leq R_e$ i.e. $n \geq 2C_1 E = 2C_1 \lceil \log(1/\sigma) \rceil$, which is true since $\sigma \succ \exp\{-n/2C_1\}$. However, the algorithm cannot stay in this phase of $\sigma \preceq R_e/n$ this until the last epoch since $\sigma > R_{E+1} = R_E/2$.

Phase 2 ($k=1$)**.** When $\sigma \succeq R_e/n$, WIDEHIST gets an estimation error of $C_2\sqrt{\frac{R_e \sigma}{n/E}}$ in epoch e. This error is the distance to the best point in epoch e, which is t by the following similar induction. In epoch e, t is still the best point only if $|x_{e-1}-t| \leq R_e$, i.e. $C_2^2 \frac{R_{e-1}\sigma}{n/E} \leq R_e^2$ i.e. $nR_e \geq 2C_2^2 E\sigma$ which holds since $R_e > \sigma$ for all $e \leq E$ and since $n \geq 2C_2^2 E$ ($\sigma \succ \exp\{-n/2C_2^2\}$ implies $E \leq n/2C_2^2$).

The final error of the algorithm is is $\sqrt{\frac{R_E \sigma}{n/E}} = \tilde{O}(\frac{\sigma}{\sqrt{n}})$ since $R_E < 2\sigma$.

Explanation for $k > 1$ Assume $\sigma \succ n^{-\frac{1}{2k-2}}$, otherwise active learning won't notice the feature noise, and so $\log(1/\sigma) \leq \frac{\log n}{(2k-2)}$. Choose total epochs $E = \lceil \log(\frac{1}{\sigma}) \rceil \leq \frac{\log n}{(2k-2)} \leq C \log n$ for some C. In each epoch of length n/E in a region of radius $R_e = 2^{-e+1}$, we get a passive bound of $C_1\sqrt{\frac{R_e}{\sigma^{2k-3} n/E}}$ whenever[1] $\sigma > (\frac{R_e}{n})^{\frac{1}{2k-1}}$.

By the same logic as for $k=1$, we need to verify that $|x_{e-1}-t| \leq R_e$ so that if t was in the search space in epoch $e-1$ then it remains the in the search space in epoch e, i.e. we want to verify $C_1^2 \frac{R_{e-1}}{\sigma^{2k-3} n/E} \leq R_e^2 \Leftrightarrow \sigma^{2k-2} R_e \geq \frac{2C_1^2 E}{n}\sigma$ which is true since[2] $R_e \geq \sigma$ and[3] $\sigma^{2k-2} > 2C_1^2 E/n$.

The final point error is given by the passive algorithm in the last epoch as $\sqrt{\frac{R_E}{\sigma^{2k-3} n/E}}$; since $R_E < 2\sigma$ and $E \leq C \log n$, this becomes $\preceq \frac{1}{\sigma^{k-2}}\sqrt{\frac{1}{n}}$.

[1] This must happen at some $e \leq E = \lceil \log(\frac{1}{\sigma}) \rceil$ because $R_E = 2^{-E+1} < 2\sigma < \sigma\sigma^{2k-2} n$ since $\sigma \succ n^{-\frac{1}{2k-2}}$ and hence in the last epoch $\sigma > (\frac{R_E}{n})^{\frac{1}{2k-1}}$.

[2] By choice of $E = \lceil \log(\frac{1}{\sigma}) \rceil$, $R_e \geq R_E \geq \sigma \geq R_{E+1}$.

[3] Since $\sigma \succ n^{-\frac{1}{2k-2}}$ we get $\sigma^{2k-2} > 2C_1^2 E/n$ since $E \leq C \log n$.