

---

# High Dimensional Statistics

## Homework 3

Javad Hezareh (404208723)

Fall 2025 (1404)

---

### 1 Mixture distributions and KL divergence

Let  $p_j$ ,  $q$  and  $\bar{q}$  be the probability density functions of  $P_j$ ,  $Q$  and  $\bar{Q}$  respectively. Then, for the LHS we have:

$$\frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || \bar{Q}) = \frac{1}{M} \sum_{j=1}^M \int p_j(x) \log \frac{p_j(x)}{\bar{q}(x)} dx$$

And for the RHS we have:

$$\frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || Q) = \frac{1}{M} \sum_{j=1}^M \int p_j(x) \log \frac{p_j(x)}{q(x)} dx$$

Now for the difference between the two sides we have:

$$\begin{aligned} & \frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || \bar{Q}) - \frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || Q) \\ &= \frac{1}{M} \sum_{j=1}^M \int p_j(x) \log \frac{p_j(x)}{\bar{q}(x)} dx - \frac{1}{M} \sum_{j=1}^M \int p_j(x) \log \frac{p_j(x)}{q(x)} dx \\ &= \frac{1}{M} \sum_{j=1}^M \int p_j(x) \left( \log \frac{p_j(x)}{\bar{q}(x)} - \log \frac{p_j(x)}{q(x)} \right) dx \\ &= \frac{1}{M} \sum_{j=1}^M \int p_j(x) \log \frac{q(x)}{\bar{q}(x)} dx \\ &= \int \left( \frac{1}{M} \sum_{j=1}^M p_j(x) \right) \log \frac{q(x)}{\bar{q}(x)} dx \\ &= \int \bar{q}(x) \log \frac{q(x)}{\bar{q}(x)} dx \\ &= -D_{KL}(\bar{Q} || Q) \end{aligned}$$

Therefore we have:

$$\frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || \bar{Q}) + D_{KL}(\bar{Q} || Q) = \frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || Q)$$

And since KL divergence is always non-negative, we have:

$$\frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || \bar{Q}) \leq \frac{1}{M} \sum_{j=1}^M D_{KL}(P_j || Q)$$

## 2 $f$ -divergences

(a) For  $f(t) = t \log t$ , we have:

$$\begin{aligned}
 D_f(P||Q) &= \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx \\
 &= \int q(x) \left(\frac{p(x)}{q(x)} \log \frac{p(x)}{q(x)}\right) dx \\
 &= \int p(x) \log \frac{p(x)}{q(x)} dx \\
 &= D_{KL}(P||Q)
 \end{aligned}$$

(b) For  $f(t) = -\log t$ , we have:

$$\begin{aligned}
 D_f(P||Q) &= \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx \\
 &= \int q(x) \left(-\log \frac{p(x)}{q(x)}\right) dx \\
 &= \int q(x) \log \frac{q(x)}{p(x)} dx \\
 &= D_{KL}(Q||P)
 \end{aligned}$$

(c) Part (d) shows that by choosing  $f(t) = 1 - \sqrt{t}$  we can get Hellinger distance.

(d) For  $f(t) = 1 - \sqrt{t}$ , we have: (We define Hellinger distance as  $H^2(P||Q) = \frac{1}{2} \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx$ )

$$\begin{aligned}
 D_f(P||Q) &= \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx \\
 &= \int q(x) \left(1 - \sqrt{\frac{p(x)}{q(x)}}\right) dx \\
 &= \int q(x) dx - \int \sqrt{p(x)q(x)} dx \\
 &= 1 - \int \sqrt{p(x)q(x)} dx \\
 &= H^2(P||Q)
 \end{aligned}$$

## 3

We know maximum is greater than or equal to the average, therefore we have:

$$\max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geq \frac{1}{2} \sum_{j=0}^1 \mathbb{P}_j(\psi \neq j)$$

We can write the RHS as follows:

$$\frac{1}{2} \sum_{j=0}^1 \mathbb{P}_j(\psi \neq j) = \mathbb{Q}(\psi \neq J)$$

Where  $\mathbb{Q}$  is the measure determining the experiment of choosing a random  $j$  uniformly from  $\{0, 1\}$  and then sampling from  $\mathbb{P}_j$ . Therefore, we have:

$$\max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geq \mathbb{Q}(\psi \neq J) \geq \inf_{\psi} \mathbb{Q}(\psi \neq J)$$

From course notes we have a lower bound for the infimal term as follows:

$$\inf_{\psi} \mathbb{Q}(\psi \neq J) \geq \frac{1}{2} (1 - \|P_0 - P_1\|_{TV})$$

Now we need to upper bound the total variation distance using KL divergence. From the definition of TV distance we have:

$$\begin{aligned} \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}^2 &= \frac{1}{4} \left( \int |p_0 - p_1| dx \right)^2 \\ &= \frac{1}{4} \left( \int |\sqrt{p_0} - \sqrt{p_1}| \cdot |\sqrt{p_0} + \sqrt{p_1}| dx \right)^2 \\ \frac{1}{4} &\leq \left( \int (\sqrt{p_0} - \sqrt{p_1})^2 dx \right) \cdot \left( \int (\sqrt{p_0} + \sqrt{p_1})^2 dx \right) \\ &= \frac{1}{4} \left( 2 - 2 \int \sqrt{p_0 p_1} dx \right) \cdot \left( 2 + 2 \int \sqrt{p_0 p_1} dx \right) \\ &= \left( 1 - \int \sqrt{p_0 p_1} dx \right) \cdot \left( 1 + \int \sqrt{p_0 p_1} dx \right) \\ &= 1 - \left( \int \sqrt{p_0 p_1} dx \right)^2 \end{aligned}$$

For the integral term we have:

$$\int \sqrt{p_0 p_1} dx = \int p_0(x) \sqrt{\frac{p_1(x)}{p_0(x)}} dx = \mathbb{E}_{P_0} \left[ \sqrt{\frac{p_1(X)}{p_0(X)}} \right] = \exp \left[ \log \left( \mathbb{E}_{P_0} \left[ \sqrt{\frac{p_1(X)}{p_0(X)}} \right] \right) \right]$$

Using Jensen's inequality we have:

$$\log \left( \mathbb{E}_{P_0} \left[ \sqrt{\frac{p_1(X)}{p_0(X)}} \right] \right) \geq \mathbb{E}_{P_0} \left[ \log \sqrt{\frac{p_1(X)}{p_0(X)}} \right] = \mathbb{E}_{P_0} \left[ \frac{1}{2} \log \left( \frac{p_1(X)}{p_0(X)} \right) \right] = -\frac{1}{2} D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)$$

Thus we have:

$$\begin{aligned} \int \sqrt{p_0 p_1} dx &\geq \exp\left(-\frac{1}{2} D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)\right) \\ \Rightarrow \left(\int \sqrt{p_0 p_1} dx\right)^2 &\geq \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)) \end{aligned}$$

Putting it all together we have:

$$\begin{aligned} \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}^2 &\leq 1 - \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)) \\ \Rightarrow 1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}^2 &\geq \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)) \\ \Rightarrow (1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV})(1 + \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}) &\geq \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)) \\ \Rightarrow 1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} &\geq \frac{\exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1))}{1 + \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}} \\ \Rightarrow 1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} &\geq \frac{\exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1))}{2} \quad (\text{since TV distance is at most 1}) \\ \Rightarrow \frac{1}{2}(1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}) &\geq \frac{1}{4} \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)) \end{aligned}$$

Therefore we have:

$$\max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geq \frac{1}{4} \exp(-D_{KL}(\mathbb{P}_0 || \mathbb{P}_1))$$

## 4 Binary Testing and Le Cam Two-Point Method

1. We know the KL divergence between two normal distributions  $N(\mu, \sigma^2)$  and  $N(\nu, \sigma^2)$  is given by:

$$D_{KL}(N(\mu, \sigma^2) || N(\nu, \sigma^2)) = \frac{(\mu - \nu)^2}{2\sigma^2}$$

From this formula it is clear that the KL divergence for two normal distributions is symmetric. Thus for the densities of this problem we have:

$$D_{KL}(N(\delta, 1) || N(-\delta, 1)) = \frac{(\delta - (-\delta))^2}{2} = \frac{(2\delta)^2}{2} = 2\delta^2$$

We also know the KL divergence for  $n$  i.i.d. samples is  $n$  times the KL divergence for one sample. Therefore we have:

$$D_{KL}(\mathbb{P}_0 || \mathbb{P}_1) = n \cdot D_{KL}(N(\delta, 1) || N(-\delta, 1)) = n \cdot 2\delta^2 = 2n\delta^2$$

2. Plugging the KL divergence into the Pinsker's inequality we have:

$$\|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} \leq \sqrt{\frac{1}{2} D_{KL}(\mathbb{P}_0 || \mathbb{P}_1)} = \sqrt{\frac{1}{2} \cdot 2n\delta^2} = \sqrt{n}\delta$$

3. Choosing  $\delta \leq \frac{1}{2\sqrt{n}}$  and plugging the bound from part 2 into Le Cam's inequality we have:

$$\begin{aligned} \inf_{\hat{\theta}} \sup_{\theta \in \{-\delta, +\delta\}} \mathbb{E}_{\theta} [(\hat{\theta} - \theta)^2] &\geq \frac{(2\delta)^2}{4} (1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}) \\ &\geq \delta^2 (1 - \sqrt{n}\delta) \\ &\geq \frac{1}{4n} \left( 1 - \sqrt{n} \cdot \frac{1}{2\sqrt{n}} \right) = \frac{1}{8n} \end{aligned}$$

## 5 Divergences and Distances between Probability Measures

(a) (a) Using the definition of Hellinger distance we have:

$$\begin{aligned}
 H^2(P||Q) &= \frac{1}{2} \sum_x \left( \sqrt{P(x)} - \sqrt{Q(x)} \right)^2 \\
 &= \frac{1}{2} \sum_x \left( P(x) + Q(x) - 2\sqrt{P(x)Q(x)} \right) \\
 &= \frac{1}{2} \left( \sum_x P(x) + \sum_x Q(x) - 2 \sum_x \sqrt{P(x)Q(x)} \right) \\
 &= 1 - \sum_x \sqrt{P(x)Q(x)}
 \end{aligned}$$

(b) First we show the lower bound. Using the definition of TV distance we have:

$$\begin{aligned}
 \|P - Q\|_{TV} &= \frac{1}{2} \sum_x |P(x) - Q(x)| \\
 &= \frac{1}{2} \sum_x |\sqrt{P(x)} - \sqrt{Q(x)}| \cdot |\sqrt{P(x)} + \sqrt{Q(x)}| \\
 &\geq \frac{1}{2} \sum_x \left( \sqrt{P(x)} - \sqrt{Q(x)} \right)^2 \quad (a, b \geq 0 \implies |a + b| \geq |a - b|) \\
 &= H^2(P||Q)
 \end{aligned}$$

And to show upper bound we have:

$$\begin{aligned}
 \|P - Q\|_{TV} &= \frac{1}{2} \sum_x |P(x) - Q(x)| \\
 &= \frac{1}{2} \sum_x |\sqrt{P(x)} - \sqrt{Q(x)}| \cdot |\sqrt{P(x)} + \sqrt{Q(x)}| \\
 &\leq \frac{1}{2} \left( \sum_x \left( \sqrt{P(x)} - \sqrt{Q(x)} \right)^2 \right)^{1/2} \cdot \left( \sum_x \left( \sqrt{P(x)} + \sqrt{Q(x)} \right)^2 \right)^{1/2} \\
 &= \frac{1}{2} (2H^2(P||Q))^{1/2} \cdot \left( 2 + 2 \sum_x \sqrt{P(x)Q(x)} \right)^{1/2} \\
 &= H(P||Q) \left( 1 + \sum_x \sqrt{P(x)Q(x)} \right)^{1/2} \\
 &\leq H(P||Q) \left( 1 + \left[ \sum_x \sqrt{P(x)}^2 \right]^{1/2} \left[ \sum_x \sqrt{Q(x)}^2 \right]^{1/2} \right) \\
 &= H(P||Q) (1 + 1 \cdot 1)^{1/2} = \sqrt{2} H(P||Q)
 \end{aligned}$$

- (b) (a) We use the definitions and Cauchy-Schwarz inequality to write:

$$\begin{aligned}
 \|P - Q\|_{TV}^2 &= \frac{1}{4} \left( \sum_x |P(x) - Q(x)| \right)^2 \\
 &= \frac{1}{4} \left( \sum_x |P(x) - Q(x)| \cdot \frac{\sqrt{Q(x)}}{\sqrt{Q(x)}} \right)^2 \\
 &\leq \frac{1}{4} \left( \sum_x \frac{(P(x) - Q(x))^2}{Q(x)} \right) \cdot \left( \sum_x Q(x) \right) \\
 &= \frac{1}{4} \left( \sum_x \frac{(P(x) - Q(x))^2}{Q(x)} \right) \cdot 1 \\
 &= \frac{1}{4} \chi^2(P||Q)
 \end{aligned}$$

(b)

- (c) (a) The given inequality does not hold in general. For  $u = 1/2 > 0$  we have:

$$\log(u) \approx -0.6931 \not\leq (u - 1) - \frac{(u - 1)^2}{2u} = -0.75$$

(b)

## 6 Le Cam's Inequality

We start with the definition of TV distance:

$$\begin{aligned}
 \|\mathbb{P} - \mathbb{Q}\|_{TV} &= \frac{1}{2} \int |p(x) - q(x)| dx \\
 &= \frac{1}{2} \int |\sqrt{p(x)} - \sqrt{q(x)}| \cdot |\sqrt{p(x)} + \sqrt{q(x)}| dx \\
 &\leq \frac{1}{2} \left( \int (\sqrt{p(x)} - \sqrt{q(x)})^2 dx \right)^{1/2} \cdot \left( \int (\sqrt{p(x)} + \sqrt{q(x)})^2 dx \right)^{1/2} \\
 &= \frac{1}{2} \left( 2 - 2 \int \sqrt{p(x)q(x)} dx \right)^{1/2} \cdot \left( 2 + 2 \int \sqrt{p(x)q(x)} dx \right)^{1/2} \\
 &= \left( 1 - \int \sqrt{p(x)q(x)} dx \right)^{1/2} \cdot \left( 1 + \int \sqrt{p(x)q(x)} dx \right)^{1/2}
 \end{aligned}$$

From the definition of Hellinger distance we have:

$$H^2(\mathbb{P}||\mathbb{Q}) = 2 - 2 \int \sqrt{p(x)q(x)} dx \implies \int \sqrt{p(x)q(x)} dx = 1 - \frac{H^2(\mathbb{P}||\mathbb{Q})}{2}$$

Thus plugging this into the previous equation we have:

$$\begin{aligned}
 \|\mathbb{P} - \mathbb{Q}\|_{TV} &\leq \left( \frac{H^2(\mathbb{P}||\mathbb{Q})}{2} \right)^{1/2} \cdot \left( 2 - \frac{H^2(\mathbb{P}||\mathbb{Q})}{2} \right)^{1/2} \\
 &= H(\mathbb{P}||\mathbb{Q}) \cdot \left( 1 - \frac{H^2(\mathbb{P}||\mathbb{Q})}{4} \right)^{1/2}
 \end{aligned}$$

## 7 Bounds for Gaussian location family

- (a) We consider  $\Theta = \{0, 2\delta\}$  as our packing. Now for  $\Phi(\cdot)$  using the two-point form of Le Cam's method we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [\Phi(\hat{\theta} - \theta)] \geq \frac{\Phi(\delta)}{2} (1 - \|\mathbb{P}_0 - \mathbb{P}_{2\delta}\|_{TV})$$

Calculating total variation distance is a little complex for this case, therefore we use an upper bound on this distance. From course notes we know:

$$\|\mathbb{P}_0 - \mathbb{P}_{2\delta}\|_{TV}^2 \leq \frac{1}{4} \left( \int \frac{p_0(x)^2}{p_{2\delta}(x)} dx - 1 \right)$$

Calculating this integral for two normal distributions we have:

$$\begin{aligned} \int \frac{p_0(x)^2}{p_{2\delta}(x)} dx &= \int \frac{\left( \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \right)^2}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-2\delta)^2}{2\sigma^2}}} dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} \frac{e^{-\frac{x^2}{\sigma^2}}}{e^{-\frac{(x-2\delta)^2}{2\sigma^2}}} dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\left( \frac{x^2}{\sigma^2} - \frac{(x-2\delta)^2}{2\sigma^2} \right)} dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\left( \frac{x^2}{2\sigma^2} + \frac{2\delta x}{\sigma^2} - \frac{2\delta^2}{\sigma^2} \right)} dx \\ &= e^{\frac{4\delta^2}{\sigma^2}} \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+2\delta)^2}{2\sigma^2}} dx \\ &= e^{\frac{4\delta^2}{\sigma^2}} \cdot 1 \\ &= e^{\frac{4\delta^2}{\sigma^2}} \end{aligned}$$

This was for single sample, for  $n$  i.i.d. samples one can simply show:

$$\int \frac{p_0^n(x)^2}{p_{2\delta}^n(x)} dx = \left( \int \frac{p_0(x)^2}{p_{2\delta}(x)} dx \right)^n = \left( e^{\frac{4\delta^2}{\sigma^2}} \right)^n = e^{\frac{4n\delta^2}{\sigma^2}}$$

Thus for the TV distance we have:

$$\|\mathbb{P}_0 - \mathbb{P}_{2\delta}\|_{TV} \leq \frac{1}{2} \sqrt{e^{\frac{4n\delta^2}{\sigma^2}} - 1}$$

Therefore we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [\Phi(\hat{\theta} - \theta)] \geq \frac{\Phi(\delta)}{2} \left( 1 - \frac{1}{2} \sqrt{e^{\frac{4n\delta^2}{\sigma^2}} - 1} \right)$$

This bound holds for any  $\delta > 0$ . Choosing  $\delta = \frac{\sigma}{2\sqrt{n}}$  we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [\Phi(\hat{\theta} - \theta)] \geq \frac{\Phi\left(\frac{\sigma}{2\sqrt{n}}\right)}{2} \underbrace{\left( 1 - \frac{1}{2} \sqrt{e - 1} \right)}_{\approx 0.345 \geq 1/3} \geq \frac{1}{6} \Phi\left(\frac{\sigma}{2\sqrt{n}}\right)$$

Now for different error functions we have:

$$\begin{cases} \Phi(\cdot) = |\cdot| \implies \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ |\hat{\theta} - \theta| \right] \geq \frac{1}{12} \frac{\sigma}{\sqrt{n}} \\ \Phi(\cdot) = (\cdot)^2 \implies \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ (\hat{\theta} - \theta)^2 \right] \geq \frac{1}{24} \frac{\sigma^2}{n} \end{cases}$$

(b) Now using Pinsker's inequality and decomposibility of KL divergence we have:

$$\begin{aligned} \|\mathbb{P}_0 - \mathbb{P}_{2\delta}\|_{TV} &\leq \sqrt{\frac{1}{2} D_{KL}(\mathbb{P}_0 \| \mathbb{P}_{2\delta})} \\ &= \sqrt{\frac{1}{2} \frac{(0 - 2\delta)^2}{2\sigma^2} \cdot n} = \sqrt{\frac{n\delta^2}{\sigma^2}} \end{aligned}$$

Plugging this into the Le Cam bound we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ \Phi(\hat{\theta} - \theta) \right] \geq \frac{\Phi(\delta)}{2} \left( 1 - \sqrt{\frac{n\delta^2}{\sigma^2}} \right)$$

This bound holds for any  $\delta > 0$  such that  $\sqrt{\frac{n\delta^2}{\sigma^2}} \leq 1$ . Choosing  $\delta = \frac{\sigma}{2\sqrt{n}}$  we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ \Phi(\hat{\theta} - \theta) \right] \geq \frac{\Phi\left(\frac{\sigma}{2\sqrt{n}}\right)}{2} \left( 1 - \frac{1}{2} \right) = \frac{1}{4} \Phi\left(\frac{\sigma}{2\sqrt{n}}\right)$$

Now for different error functions we have:

$$\begin{cases} \Phi(\cdot) = |\cdot| \implies \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ |\hat{\theta} - \theta| \right] \geq \frac{1}{8} \frac{\sigma}{\sqrt{n}} \\ \Phi(\cdot) = (\cdot)^2 \implies \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \left[ (\hat{\theta} - \theta)^2 \right] \geq \frac{1}{16} \frac{\sigma^2}{n} \end{cases}$$

## 8 Normal location model via Fano method

Using Fano's method we can construct a packing as  $\Theta = \{0, 2\delta, -2\delta\}$ . Now writing the Fano's inequality for  $\Phi(\cdot) = (\cdot)^2$  we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [(\hat{\theta} - \theta)^2] \geq \delta^2 \left( 1 - \frac{\frac{1}{|\Theta|^2} \sum_{i,j} D_{KL}(\mathbb{P}_{\theta_i} || \mathbb{P}_{\theta_j}) + \log 2}{\log |\Theta|} \right)$$

For the KL divergences we have:

$$\begin{cases} \theta_i = \theta_j & \implies D_{KL}(\theta_i, \theta_j) = 0 \\ \theta_i = 0, \theta_j = \pm 2\delta & \implies D_{KL}(\theta_i, \theta_j) = \frac{(0 - (\pm 2\delta))^2}{2\sigma^2} \cdot n = \frac{2n\delta^2}{\sigma^2} \\ \theta_i = 2\delta, \theta_j = -2\delta & \implies D_{KL}(\theta_i, \theta_j) = \frac{(2\delta - (-2\delta))^2}{2\sigma^2} \cdot n = \frac{8n\delta^2}{\sigma^2} \end{cases}$$

Thus for the average KL divergence term we have:

$$\begin{aligned} \frac{1}{|\Theta|^2} \sum_{i,j} D_{KL}(\mathbb{P}_{\theta_i} || \mathbb{P}_{\theta_j}) &= \frac{1}{9} \left( 0 + 2 \cdot \frac{2n\delta^2}{\sigma^2} + 2 \cdot \frac{2n\delta^2}{\sigma^2} + 2 \cdot \frac{8n\delta^2}{\sigma^2} \right) \\ &= \frac{8n\delta^2}{3\sigma^2} \end{aligned}$$

Plugging this into the Fano's inequality we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [(\hat{\theta} - \theta)^2] \geq \delta^2 \left( 1 - \frac{\frac{8n\delta^2}{3\sigma^2} + \log 2}{\log 3} \right)$$

This bound holds for any  $\delta > 0$  such that  $\frac{8n\delta^2}{3\sigma^2} + \log 2 \leq \log 3$ . Choosing  $\delta = \frac{\sigma}{2\sqrt{n}}$  we have:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} [(\hat{\theta} - \theta)^2] \geq \frac{1}{4} \frac{\sigma^2}{n} \left( 1 - \frac{\frac{8}{3} \cdot \frac{1}{4} + \log 2}{\log 3} \right) = \frac{1}{4} \frac{\sigma^2}{n} \underbrace{\left( 1 - \frac{2/3 + \log 2}{\log 3} \right)}$$

## 9 Minimax risk for sparse linear regression