Solution to PSET 3

- **Do not** distribute the solutions outside the class.

1. **(Minimax Lower Bounds for the Uniform Location Family)**
(a) We have

$$\begin{aligned}\mathbb{E}_\theta[(X_{(1)}-\theta)^2] &= \int_0^1 \mathbb{P}_\theta((X_{(1)}-\theta)^2 \geq t) \\ &= \int_0^1 \mathbb{P}_\theta((X_{(1)}-\theta) \geq \sqrt{t}) \\ &= \int_0^1 (1-\sqrt{t})^n dt \\ &= \frac{2}{(n+1)(n+2)}.\end{aligned}$$

(b) To apply the Le Cam's two point method for deriving a lower bound on Minimax risk, consider two uniform distributions with parameters $\theta_1 = 0$ and $\theta_2 = 2\delta, \delta < \frac{1}{2}$. Then the minimax risk is lower-bounded by the following quantity

$$\frac{1}{2}\delta^2(1-||P_1^n - P_2^n||). \tag{1}$$

We now compute

$$d^2_{\text{hel}}(P_1, P_2) = 4\delta.$$

Hence,

$$d^2_{\text{hel}}(P_1^n, P_2^n) = 2\bigg(1-(1-2\delta)^n\bigg).$$

Thus,

$$\begin{aligned}||P_1^n - P_2^n|| &\leq \sqrt{2\big(1-(1-2\delta)^n\big)}\sqrt{1-\frac{1}{2}(1-\big(1-2\delta)^n\big)} \\ &= \sqrt{1-(1-2\delta)^{2n}} \\ &\leq 1-\frac{1}{2}(1-2\delta)^{2n}.\end{aligned}$$

Hence, from Eqn. (1) the minimax risk is lower bounded by

$$\frac{1}{4}\delta^2(1-2\delta)^{2n} \geq \frac{1}{4}\delta^2(1-4n\delta).$$

Now, letting $\delta = \frac{1}{8n}$, the minimax risk is lower bounded by $\frac{c}{n^2}$, where $c = \frac{1}{512}$.)

2. **(KL Divergence and Differential Privacy)** (a) WOLOG, assume that $a \geq b > 0$. Then

$$|\ln \frac{a}{b}| = \ln \frac{a}{b} = \ln(1 + \frac{a-b}{b}) \leq \frac{a-b}{b}.$$

(b) We have

$$\begin{aligned} & m_1(z) - m_2(z) \\ = & \int q(z|x)(p_1(x) - p_2(x))dx \\ = & \int \big(q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\big)\big(p_1(x) - p_2(x)\big)dx \end{aligned}$$

Hence, using the triangle inequality, we have

$$\begin{aligned} & |m_1(z) - m_2(z)| \\ \leq & \int |\big(q(z|x) - \inf_{x\in\mathcal{X}} q(z|x)\big)||\big(p_1(x) - p_2(x)\big)|dx \\ \leq & (e^\alpha - 1) \inf_{x\in\mathcal{X}} q(z|x)||P_1 - P_2||_{\text{TV}}. \end{aligned}$$

where the last inequality follows from the definition of differential privacy.
(c) We have

$$\begin{aligned} & D(M_1||M_2) + D(M_2||M_1) \\ \leq & \sum_z \big(m_1(z) - m_2(z)\big) \ln\left(\frac{m_1(z)}{m_2(z)}\right) \\ \stackrel{(a)}{\leq} & \sum_z \Big(m_1(z) - m_2(z)\Big)^2 \frac{1}{\min\{m_1(z), m_2(z)\}} \\ \stackrel{(b)}{\leq} & \sum_z \frac{\big(m_1(z) - m_2(z)\big)^2}{\inf_{x\in\mathcal{X}} q(z|x)} \\ \stackrel{(c)}{\leq} & (e^\alpha - 1)^2||P_1 - P_2||^2 \sum_z \inf_x q(z|x) \\ \stackrel{(d)}{\leq} & (e^\alpha - 1)^2||P_1 - P_2||^2, \end{aligned}$$

where the inequality (a) follows from part (a), the inequality (b) follows from the fact that $m_i(z) = \sum_x q(z|x)p_i(x) \geq \inf_x q(z|x)$, the inequality (c) follows from part (b), and finally, the inequality (d) follows from the fact that $\sum_z \inf_x q(z|x) \leq \sum_z q(z|x_1) = 1$.

3. **(Application of Le Cam's method to detecting drug abuse)** (a) To apply Le Cam's two point method, consider two parameters $\theta_1 = \frac{1}{2}, \theta_2 = \frac{1}{2} + 2\delta, \delta \leq \frac{1}{4}$.
We have

$$D(P_1||P_2) \leq c_1\delta^2,$$

for some numerical constant c_1. Hence,

$$D(P_1^n||P_2^n) = nD(P_1||P_2) \leq nc_1\delta^2.$$

Making use of the Strong data processing inequality and Pinsker's inequality, we have

$$D(M_1^n||M_2^n) \leq nc_2(e^\alpha - 1)^2\delta^2.$$

Using Two point tests and Pinsker's inequality once again, we now have

$$\begin{aligned} \mathcal{M}_n &\geq \frac{1}{2}\delta(1 - \sqrt{n}c_2'(e^\alpha - 1)\delta) \\ &\geq \frac{1}{2}\delta\big(1 - c_2''\alpha\sqrt{n}\delta\big). \end{aligned}$$

Taking $c_2''\sqrt{n}\alpha\delta = \frac{1}{2}$, i.e., $\delta = \frac{c_4}{\sqrt{n\alpha^2}}$, for some appropriate numerical constant c_4, we have

$$\mathcal{M}_n \geq \frac{c}{\sqrt{n\alpha^2}},$$

for some numerical constant c.

(b) Define the following channel

$$\begin{aligned} Q_\beta(Z = 0|X = 0) = Q_\beta(Z = 1|X = 1) = \frac{1}{2} + \beta, \\ Q_\beta(Z = 1|X = 0) = Q_\beta(Z = 0|X = 1) = \frac{1}{2} - \beta. \end{aligned}$$

To ensure that the channel is α-differentially private for some $\alpha \leq \frac{1}{2}$, we require the channel parameter β to satisfy:

$$\frac{\mathbb{P}(Z_i = 0|X_i = 0)}{\mathbb{P}(Z_i = 1|X_i = 0)} = \frac{1 + 2\beta}{1 - 2\beta} \leq \exp(\alpha),$$

i.e., $\beta \leq \frac{1}{2}\frac{e^\alpha - 1}{e^\alpha + 1}$. Since $\frac{e^\alpha - 1}{e^\alpha + 1} \geq \frac{\alpha}{3}, 0 \leq \alpha \leq \frac{1}{2}$. It suffices to take $\beta = \alpha/6$.
Now,

$$\begin{aligned} \mathbb{E}_\theta(Z_i) &= 1/2 - \beta + 2\beta\theta, \quad \forall i. \\ \mathrm{Var}(Z_i) &\leq \frac{1}{4}. \end{aligned}$$

Next, consider the following estimator for θ:

$$\hat{\theta}(Z^n) = \frac{1}{2\beta}\big(\frac{1}{n}\sum_{i=1}^n Z_i - (1/2 - \beta)\big). \tag{2}$$

Hence, $\mathbb{E}_\theta(\hat{\theta}(Z^n)) = \theta$. Using the Jensen's inequality, we also have

$$|\mathbb{E}_\theta(\hat{\theta}(Z^n)) - \theta| \leq \sqrt{\mathrm{Var}(\hat{\theta}(Z^n))} \leq \frac{c_1}{\sqrt{n\beta^2}} = \frac{C}{\sqrt{n\alpha^2}}, \tag{3}$$

for some numerical constant C.
(c)
(d) The result of using the estimator (2) in the given dataset is presented in the tabular form below, where we average over $N_{\text{expt}} = 100$ times to get the average accuracy figures.

α	Accuracy
2^{-1}	0.0024
2^{-2}	0.0154
2^{-3}	0.0289
2^{-4}	0.0567
2^{-5}	0.1291
2^{-6}	0.2296
2^{-7}	0.4629
2^{-8}	1.0892
2^{-9}	1.8802
2^{-10}	3.8776

Plotting the results (on a log-log scale) corroborates the dependence of the privacy parameter α on the theoretical estimation error bound given in Eqn. (3).

4. **(Fundamental Limit of Sign Identification in Sparse Signals)**

Let the signal S be chosen uniformly at random from the signal set

$$\mathcal{S}_k = \{s \in \{-1, 0, +1\}^d : ||s||_1 = k\}.$$

Simple combinatorics tells us that $|\mathcal{S}_k| = \binom{d}{k}2^k$. Using Fano's inequality, we have

$$\mathbb{P}(\hat{S} \neq S) \geq 1 - \frac{I(Y;S) + \ln 2}{\ln |\mathcal{S}_k|}.$$

Hence,

$$\mathbb{P}(\hat{S} \neq S) \geq \frac{1}{2} \quad \text{unless} \quad \frac{I(Y;S) + \ln 2}{\ln |\mathcal{S}_k|} \geq \frac{1}{2}. \tag{4}$$

In the rest of the derivations, we compute an upper bound for the mutual information $I(Y;S)$. Recall that $I(Y;S) = h(Y) - h(Y|S)$. Since $Y = X\theta^S + \epsilon$, we have

$$h(Y|S) = h(X\theta^S + \epsilon|S) = h(\epsilon) = \frac{n}{2}\ln(2\pi e\sigma^2).$$

Next, we obtain an upper bound for $h(Y)$. Note that $\mathbb{E}(Y) = \mathbf{0}$ and its covariance

$$\mathbb{E}(YY^T) = \mathbb{E}_{S,\epsilon}\bigg((X\theta^S + \epsilon)((\theta^S)^T X^T + \epsilon^T)\bigg) = X\mathbb{E}_S(\theta^S(\theta^S)^T)X^T + \sigma^2 I_n.$$

Finally, note that $\mathbb{E}_S((\theta^S)_i^2) = \frac{k}{d}\theta_{\min}^2$, and for $i \neq j$, $\mathbb{E}(\theta_i^S \theta_j^S) = 0$. Hence, the covariance matrix is simplified to

$$K_{YY} = \frac{k}{d}\theta_{\min}^2 XX^T + \sigma^2 I_n.$$

Since, jointly normal distribution maximizes entropy with a fixed covariance matrix, we have

$$h(Y) \leq \frac{1}{2}\ln(2\pi e)^n \mathsf{det}(K_{YY}).$$

Finally, we derive an upper bound for the determinant of the real symmetric PD matrix K_{YY}. The summation of n eigenvalues of the matrix K_{YY} is computed as

$$\mathsf{Tr}(K_{YY}) = \frac{k}{d}\theta_{\min}^2 ||X||_{\mathsf{Fr}}^2 + n\sigma^2.$$

Using the AM-GM inequality, the product of the eigenvalues, *i.e.*, $\mathsf{det}(K_{YY})$ is upper bounded as

$$\mathsf{det}(K_{YY}) \leq \big(n^{-1}\mathsf{Tr}(K_{YY})\big)^n = \left(\frac{k}{d}\theta_{\min}^2 ||n^{-1/2}X||_{\mathsf{Fr}}^2 + \sigma^2\right)^n.$$

Thus,

$$h(Y) \leq \frac{n}{2}\ln(2\pi e) + \frac{n}{2}\ln\left(\frac{k}{d}\theta_{\min}^2 ||n^{-1/2}X||_{\mathsf{Fr}}^2 + \sigma^2\right).$$

This gives us the following upper bound on the mutual information:

$$I(Y;S) \leq \frac{n}{2}\ln\left(\frac{k}{d}\frac{\theta_{\min}^2}{\sigma^2}||n^{-1/2}X||_{\mathsf{Fr}}^2 + 1\right) \leq \frac{n}{2}\frac{k}{d}\frac{\theta_{\min}^2}{\sigma^2}||n^{-1/2}X||_{\mathsf{Fr}}^2,$$

where, in the last inequality, we have used the fact that $\ln(1+x) \leq x, \forall x \geq 0$. Hence, Eqn. (4) implies that $\mathbb{P}(\hat{S} \neq S) \geq \frac{1}{2}$ unless

$$\frac{\frac{n}{2}\frac{k}{d}\frac{\theta_{\min}^2}{\sigma^2}||n^{-1/2}X||_{\mathsf{Fr}}^2 + \ln 2}{\ln|\mathcal{S}_k|} \geq \frac{1}{2}.$$

i.e.,

$$n \geq \frac{\frac{d}{k}\ln\binom{d}{k}}{||n^{-1/2}X||_{\mathsf{Fr}}^2}\frac{\sigma^2}{\theta_{\min}^2},$$

where we have used the fact that $k \geq 2$.
(b) Since $X \in \{-1, +1\}^{n\times d}$, we have $||n^{-1/2}X||^2_{\mathsf{Fr}} = d$. Thus, for correct recovery with probability at least $\frac{1}{2}$, we must have

$$n \geq \ln \binom{d}{k} \frac{\sigma^2}{k\theta^2_{\min}} = \frac{\ln \binom{d}{k}}{k\mathsf{SNR}},$$

where $\mathsf{SNR} \equiv \frac{\theta^2_{\min}}{\sigma^2}$, denotes the signal-to-noise ratio per received symbol.

5. **(VC-dimension of Polynomials)** (a) Note that the polynomial functions are continuous and each sign change of a polynomial function is in one-to-one correspondence of a real root of the polynomial. Since a polynomial of degree d defined over the reals can have at most d roots, it follows that $p \in \mathcal{H}_d$ can have at most d sign changes over $\mathbb{R}$.
(b) Select S to be the set of first $d+1$ integers, i.e., $S = \{1, 2, \ldots d+1\}$. Consider any labelling of S. If there is a sign change between the labellings of integers i and $i+1$ in S, add a root $\alpha = i + \frac{1}{2}$ to the polynomial $p(x)$ else, continue.
Since there can be at most d sign changes, we have at most d roots of $p(x)$ and hence $p \in \mathcal{H}_d$. It is clear that the polynomial $\pm p(x)$ produces the desired labeling of the set S. Hence VC dimension of $\mathcal{H}_d$ is at least $d+1$.
(c) Consider a set S with $|S| = d+2$. Arrange the elements of S in increasing order. Consider an alternating labeling of S. Hence we have at least $d+1$ sign changes of any polynomial p which produces the desired labeling of S. From part (a) we conclude that such a polynomial does not belong to $\mathcal{H}_d$.

6. **(VC dimension of Boolean Conjunctions)** Recall that boolean conjunctions are of the form $f(\boldsymbol{x}) = \big(\wedge_{i\in S_1} x_i\big) \wedge \big(\wedge_{j\in S_0} x_j^c\big)$, for two disjoint subsets $S_0, S_1 \subset \{1, 2, \ldots, d\}$.
(a) Consider the set of all boolean functions with k literals. We can choose the literals in $\binom{d}{k}$ ways. With each choice of k literals, depending on which variable we complement, there are 2^k possible functions. Hence,

$$|\mathcal{H}^d_{\mathsf{con}}| \leq \sum_{k=0}^{d} \binom{d}{k} 2^k = 3^d + 1.$$

(b) The above shows that $\mathsf{VCdim}(\mathcal{H}) \leq d \log 3$.

(c) For any given labelling $l : \boldsymbol{x} \to \{0, 1\}$, of the unit vectors, consider the sets $S_+ = \{i : l(\boldsymbol{e}_i) = 1\}$ and $S_- = \{i : l(\boldsymbol{e}_i) = 0\}$, we can get an equivalent label $f(\boldsymbol{x}) = \wedge_{i\in S_-} x_i^c$.