Homework 4
36-705
Due: Thursday October 3rd by 3pm

1. Recall that the Rademacher complexity for a class of functions was defined as:

\[ R(\mathcal{F}) = \mathbb{E}_{\epsilon, X} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) \epsilon_i \right| . \]

Though we did not explore this in lecture one is often (for instance when analyzing regression problems) interested in bounding the Rademacher complexity for classes of functions that are not indicators of sets. We will do this for linear regressors in this exercise, i.e. the class of functions:

\[ \mathcal{F} = \{ f : f(x) = \langle \beta, x \rangle, \| \beta \|_2 \leq B \} . \]

We will also suppose that each \( X_i \sim N(0, I_d) \). Now show that:

\[ R(\mathcal{F}) \leq B \sqrt{\frac{d}{n}}. \]

**Hint:** Depending on how you choose to solve this you might find Jensen’s inequality helpful, particularly that \( \mathbb{E}[X] \leq \sqrt{\mathbb{E}[X^2]} \).

2. Suppose that we take a collection of sets \( \mathcal{A} \), and a collection of sets \( \mathcal{B} \), and define \( \mathcal{C} \) as:

\[ \mathcal{C} = \{ A \cup B : A \in \mathcal{A}, B \in \mathcal{B} \} . \]

Show that the shattering number:

\[ s(\mathcal{C}, n) \leq s(\mathcal{A}, n) \times s(\mathcal{B}, n) . \]

3. Suppose instead of taking the union of individual sets, we simply collected all sets to define:

\[ \mathcal{C} = \{ A : A \in \mathcal{A} \text{ or } A \in \mathcal{B} \} . \]

Show that the shattering number:

\[ s(\mathcal{C}, n) \leq s(\mathcal{A}, n) + s(\mathcal{B}, n) . \]

4. Suppose that \( \theta > 0 \), let \( X_1, \ldots, X_n \sim \text{Unif}(-\theta, \theta) \):

(a) Write down the likelihood function.

(b) Find a minimal sufficient statistic.
(c) Show that $X_1$ is not a sufficient statistic.

5. Suppose that $X_1, \ldots, X_n \sim N(\mu, \mu^2)$. Compute the likelihood function. Find a minimal sufficient statistic.

6. Suppose that we have $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. We begin with the initial estimator

$$\hat{\theta} = X_1,$$

and take the sufficient statistic $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$.

(a) Compute the Rao-Blackwellized estimator.

**Hint:** Use the fact that $(X_1, \sum_{i=1}^{n} X_i)$ has a joint Gaussian distribution. Compute its mean and covariance. You can then look up on Wikipedia what the conditional distribution is for a bivariate Gaussian.

(b) Compute the risk of $\hat{\theta}$ and compare it to the risk of the Rao-Blackwellized estimator.

(c) Suppose you instead tried to compute the “estimator” $E[X_1|X_2]$. Show that this is infeasible.

7. Prove that the likelihood induces the minimal sufficient partition, i.e. suppose that we grouped sequences $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ if $L(\theta; x_1, \ldots, x_n) \propto L(\theta; y_1, \ldots, y_n)$ i.e. the likelihoods are the same (possibly up to a positive constant), then we obtain the minimal sufficient partition. Incidentally, this method does not give us a minimal sufficient statistic.

8. If $X_1, \ldots, X_n \sim U[\theta, 1 + \theta]$ then prove that

$$T(X_1, \ldots, X_n) = (\min_i X_i, \max_i X_i)$$

is a minimal sufficient statistic. (We will verify some part of this statement in class but you need to provide a full proof for the HW).