ESE 504-542 : Statistics for Data Science
Instructor: Hamed Hassani, Shirin Saeedi
Spring 2019

Final Examination

NAME	

One two-sided note-sheet allowed.

	Grade (y/n)	Score	Max. Score
Problem 1			40
Problem 2			40
Problem 3			20
TOTAL			100

Problem 1 (40 points) [**Simple Linear Regression.**]
Consider the following simple linear regression problem with the data set (x_1, y_1), $(x_2, y_2), \ldots,$ (x_n, y_n).

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \tag{1}$$

Assume all assumptions for linear regression are met. Particularly, ϵ_i are i.i.d. random variables where $\epsilon_i \sim N(0, \sigma^2)$.

1. Derive the estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ by minimizing the residual sum of squares i.e., by solving

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

2. Derive the estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ using maximum likelihood estimation i.e., by solving

$$\max_{\beta_0,\beta_1} \ \log \ell(\beta_0, \beta_1),$$

where

$$\ell(\beta_0, \beta_1) = \prod_{i=1}^{n} \Pr\left(y_i \mid \beta_0, \beta_1, x_i\right).$$

Note that $\Pr\left(y_i \mid \beta_0, \beta_1, x_i\right)$ is the probability of observing y_i given the values of β_0, β_1 and x_i. Compare the results with part 1.

3. Show that your estimates are unbiased i.e., show that

$$E\left[\hat{\beta}_0\right] = \beta_0, \qquad E\left[\hat{\beta}_1\right] = \beta_1.$$

4. Consider the case when heteroskedasticity is present, i.e., $\epsilon_i \sim N(0, \sigma_i^2)$. Repeat part 2 under heteroskedasticity.

Problem 2 (40 points) [**Weighted K-Means Clustering**.]
Consider data points $x_1, x_2, \cdots, x_n \in \mathbb{R}^d$. For each data point x_i we have assigned a positive number $w_i \geq 0$ which indicates the importance of that data point. Our goal is to provide an algorithm for the following *weighted K-Means clustering* problem: Find K centers $c_1, c_2, \cdots, c_K \in \mathbb{R}^d$ that minimize the objective

$$\sum_{i=1}^{n} w_i \times \min_{j \in \{1, \cdots, K\}} ||x_i - c_j||^2. \tag{2}$$

1. Assume that $K = 1$. Find the optimal centroid that minimizes (2).

2. For a given K, Extend the K-Means algorithm taught in class to the weighted setting. Explain precisely what the algorithm is and justify your answer

Problem 3 (20 points) [**Basic Questions about Learning Theory.**]

1. Give a precise definition of "PAC Learnability".

2. Explain briefly why finite hypothesis classes are PAC learnable.

3. What property should an infinite hypothesis class have in order to be PAC learnable?