EEE 485-585 SPRING 2018 PROBLEM SET 1

Due Date: February 26 2018, 17:30

Question 1 [25 pts]

Suppose that a student from Statistical Learning course would like to buy a comfortable keyboard online for his programming assignments. He finds out that only two retailers from the shopping website have the keyboard model he prefers. The retailers list the keyboard with the same price tag. However, the student observes that the first retailer has 19 positive and 1 negative feedbacks while the second retailer has 830 positive and 60 negative feedbacks. The student also knows that the computer hardware retailers on the website have a mean of 0.91 positive feedback rate with 0.0004 variance. The student, therefore, decides to choose the retailer according to the maximum a posteriori (MAP) estimates of the probability of good service from the retailers, θ_1 and θ_2, respectively, with a Beta prior.

(a) [4 pts] Model the prior information using a Beta distribution. Find the parameters α and β.

(b) [7 pts] Find MAP estimates for θ_1 and θ_2. According to the estimates, which retailer should the student select, the first or the second?

(c) [7 pts] Which retailer would the student select if he, instead, decides to use the maximum likelihood estimates (MLE) of θ_1 and θ_2?

(d) [7 pts] Using the posterior distributions, find the central credible interval ($\alpha = 0.1$) for θ_1 and θ_2. Which credible interval is wider? Based on the credible intervals, which retailer shold the student select, the first or the second?

Question 2 [20 pts]

(a)[6 pts] Let $Y_1, Y_2, \ldots, Y_n$ (take n as $100, 1000, 1000$) be an i.i.d. sequence of normal random variables with mean $\mu = 0.5$ and standard deviation 0.3. Form the sample mean given as $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ and the sample variance given as $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \hat{\mu})^2$. Repeat this experiment 100 times and (i) plot the 100 sample mean values you obtained (ii) 100 sample variance values you obtained. What do you observe as n increases?

(b)[7 pts] What is the distribution of $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. What is its expected value? What is its variance?

(c)[7 pts] What is the distribution of $\hat{\sigma}^2$? What is its expected value? What is its variance?

Question 3 [15 pts]

(b)[7 pts] Consider n data samples $y_1, \ldots, y_n$ drawn independently from normal distribution $\mathcal{N}(\mu_{\text{true}}, \sigma^2_{\text{true}})$. Calculate the maximum likelihood estimates $\hat{\mu}_{\text{MLE}}$ and $\hat{\sigma}^2_{\text{MLE}}$.

(b)[8 pts] Calculate the expected values of $\hat{\mu}_{\text{MLE}}$ and $\hat{\sigma}^2_{\text{MLE}}$. Compare them with the expected values of the sample mean and the sample variance found in Question 2. Are they the same?

Question 4 [15 pts]

(a)[12 pts] Download the PS1train and PS1test sets from Moodle. First column is x and second column is y. Train p different polynomial regression models $p = 1, 2, \ldots, 10$ on the training set. On the same figure, report both the train and test RSS as a function of p. Based on your findings, which model will you choose? Explain clearly. Include your code with the submission. Using machine learning libraries is not allowed.

(b)[3 pts] Using what you have found above, predict the output when $x = 5$.

Question 5 [10 pts]

Suppose we have a dataset that contains the financial status of startups that have recently entered the stock market with three regressors, X_1 = "raised funds" (in millions of dollars), X_2 = "initial stock value", X_3 = "debt" (in millions of dollars). The variable of interest Y is the company value after a year (in millions of dollars). Suppose we use least squares to fit the model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 \ln X_{i2} + X_{i3}$ and get $\beta_0 = 10$, $\beta_1 = 10$, $\beta_2 = 0.5$, $\beta_3 = -5$.

Which answer is correct, and why?

1. One unit change X_1 causes a 1000 percent change in Y.
2. One unit change in X_2 causes a 50 percent change in Y.
3. 100 percent change in X_2 causes a 50 percent change in Y.
4. Higher debt implies lower future stock value.
5. Is there any meaningful interpretation of the bias term β_0?

Question 6 [15 pts]

(a) [3 pts] You are given a dataset that is composed of instances $\{x_t\}_{t=1}^{T}$, $x_t \in \mathbb{R}$ where the true relation between x_ts is given by

$$x_t = 0.9x_{t-1} - 0.6\sqrt{x_{t-2}} + 2$$

The prediction for the tth instance $t > 3$ is given by

$$\hat{x}_t = ax_{t-1} + b\sqrt{x_{t-2}} + cx_{t-3} + d.$$

Find a, b, c, d, in terms of $\{x_t\}_{t=1}^{T}$ such that $\sum_{t=4}^{T}(\hat{x}_t - x_t)^2$ is minimized.

(b) [3 pts] Fix $T = 30$, $x_2 = 1$ and $x_1 = 0$. Generate the dataset $\{x_t\}_{t=1}^{T}$. Using the method you proposed above, find the values of the coefficients a, b, c, d and the value of $\sum_{t=4}^{T}(\hat{x}_t - x_t)^2$. Using machine learning libraries is not allowed.

(c) [9 pts] Use the same dataset in part b. Propose a method to minimize $\sum_{t=4}^{T}|\hat{x}_t - x_t|$. Find the values of the corresponding coefficients and the value of $\sum_{t=4}^{T}|\hat{x}_t - x_t|$.