CS5691: Pattern recognition and machine learning
Quiz - 2
Course Instructor : Prashanth L. A.
Date : Feb-22, 2019 **Duration** : 40 minutes

Name of the student :
Roll No :

INSTRUCTIONS: For MCQ questions, you do not have to justify the answer. For the rest, provide proper justification for the answers. Please use rough sheets for any calculations *if necessary.* Please **DO NOT** submit the rough sheets. DO NOT use pencil for writing the answers.

I. Multiple Choice Questions

Note: 1 mark for the correct answer. Only one answer is correct. Please write the choice code a, b, c or d in the answer box provided.

(1) Let $\{X_1, \ldots, X_n\}$ be i.i.d. samples from $\mathbb{N}(\mu, \sigma^2)$, with $\sigma > 0$. Letting $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then, which of the following statements is true?

(a) $\sum_{i=1}^n (X_i - \hat{\mu}_n)^2 = \sum_{i=1}^n (X_i - \mu)^2$.
(b) $\sum_{i=1}^n (X_i - \hat{\mu}_n)^2 \leq \sum_{i=1}^n (X_i - \mu)^2$.
(c) $\sum_{i=1}^n (X_i - \hat{\mu}_n)^2 > \sum_{i=1}^n (X_i - \mu)^2$.
(d) An inequality/equality relating $\sum_{i=1}^n (X_i - \hat{\mu}_n)^2$ and $\sum_{i=1}^n (X_i - \mu)^2$ does not always hold.

Answer: ☐

(2) Consider a Bayesian estimation problem, with data $\{X_1, \ldots, X_n\}$ i.i.d. from $\mathbb{N}(\theta, 1)$, and a $\mathbb{N}(0, 1)$ prior. Letting $S_n = \sum_{i=1}^n X_i$, the posterior mean is

(a) $\dfrac{S_n}{n}$
(b) $\dfrac{S_n}{n+1}$
(c) $\dfrac{nS_n}{n+1}$
(d) $\dfrac{S_n+1}{n+2}$

Answer: ☐

(3) Let $X \sim \text{Unif}[0, \theta]$. Then, the maximum likelihood estimate of θ, given i.i.d. samples $\{X_1, \ldots, X_n\}$ is

(a) $\sum_{i=1}^n \frac{S_n}{n}$.
(b) $\min_{i=1,\ldots,n} X_i$.
(c) $\max_{i=1,\ldots,n} X_i$.
(d) $\frac{1}{2}\left(\max_{i=1,\ldots,n} X_i - \min_{i=1,\ldots,n} X_i\right)$.

Answer: ☐

(4) Suppose that we are trying to fit a linear and 10th degree polynomial to data coming from a cubic function, corrupted by standard Gaussian noise. Let M_1 and M_2 denote the models corresponding to the linear and 10 degree polynomial. Then,

(a) $\text{Bias}(M_1) \leq \text{Bias}(M_2)$, $\text{Variance}(M_1) \leq \text{Variance}(M_2)$.
(b) $\text{Bias}(M_1) \leq \text{Bias}(M_2)$, $\text{Variance}(M_1) \geq \text{Variance}(M_2)$.
(c) $\text{Bias}(M_1) \geq \text{Bias}(M_2)$, $\text{Variance}(M_1) \leq \text{Variance}(M_2)$.
(d) $\text{Bias}(M_1) \geq \text{Bias}(M_2)$, $\text{Variance}(M_1) \geq \text{Variance}(M_2)$.

Answer: ☐

(5) Consider a regression problem, with scalar input $X \in \mathbb{R}$, and target $Y \in \mathbb{R}$. Suppose (X, Y) is bivariate normal with non-zero means, positive variances, and non-zero correlation. Then, the optimal predictor, for the square loss, as a function of X is

(a) Quadratic. (b) Constant.

(c) Linear. (d) None of the above.

Answer: ☐

II. A problem that requires a detailed solution

(1) Consider a distribution over (X, Y) given by the following assumptions:
$Y \in \{-1, +1\}, X \in \{0, 1\}^3$.
$\mathbb{P}(Y = +1) = a, \mathbb{P}(Y = -1) = 1 - a$,
$X|Y = -1 \sim \text{Bern}(\theta_1) \times \text{Bern}(\theta_2) \times \text{Bern}(\theta_3)$,
$X|Y = +1 \sim \text{Bern}(\tau_1) \times \text{Bern}(\tau_2) \times \text{Bern}(\tau_3)$.

We have 10 training points from the above distribution, given by the table below.

X_1	X_2	X_3	Y
1	0	0	$+1$
0	1	1	-1
0	1	0	$+1$
1	1	0	$+1$
1	1	1	-1
1	0	0	$+1$
1	0	1	$+1$
0	0	1	-1
0	1	1	$+1$
0	0	0	-1

i. Give the ML estimates for $a, \theta_1, \theta_2, \theta_3, \tau_1, \tau_2, \tau_3$. (3 marks)

ii. For all the 8 points in the instance space $\{0, 1\}^3$, give the estimate of the posterior probability $\mathbb{P}(Y = +1 \mid X)$, and give the prediction that minimises the mis-classification rate (or the Bayes classifier for the zero-one loss), in the form of a table with 8 rows. (2 marks)