SML 2024, Monsoon, Optional quiz, Duration 45 minutes

Q1. A machine' functioning probability depends on two parameters - maintenance p, and load h. Let $p, h \in \{0, 1\}$ and are independent of each other. $p = 1$ means that machine has a good maintenance record and $p = 0$ means poor maintenance record. Similarly, for $h = 1$ means less load and 0 means overload. If a machine has all parameters as 1, then the chances of good functioning is very high. Suppose you start experimenting and come up with the data in Table 1.

Table 1: Survey response

	p	h
S-1	1	0
S-2	1	1
...	..	...
S-i	0	1

In Table 1, S-i refers to i^{th} response. There are a total of n *iid* responses.

a. Derive a general expression to determine the probability of p being 1, and h being 1, that is compute $\theta_1 = prob(p = 1)$, and $\theta_2 = prob(h = 1)$. Note the expression should be a function of n and S-i. [2]

b. Suppose we observe $p = 0, h = 1$. Use the first two rows, namely, S-1 and S-2 of Table 1 to determine θ_1 and θ_2. Compute the functioning probability of the machine. [1] [CO1]

Q2. Given classification data $D = \{(x, y)\} = \{(-1, 0), (-3, 0), (1, 0), (2, 1), (5, 1)\}$. y is the label. Using bagging three different trees are obtained $h_1(x), h_2(x), h_3(x)$. $h_1(x)$ is obtained by a split at $x = 0$, $h_2(x)$ is obtained by a split at $x = 3$, and $h_3(x)$ is obtained by a split at $x = 3.5$. Find the class of point $x = 1.5$ using the bagged tree. [2] [CO2]

Q3. a. Determine β^* corresponding to Chernoff bound for two category case where both the categories follow a Gaussian distribution. Both categories have same mean. Variance of category 1 is 1. Second category has variance of 2. Assume equal priors. Hint: [2] [CO1]

$k(\beta) = \frac{\beta(1-\beta)}{2}(\mu_2 - \mu_1)^\top [\beta\Sigma_1 + (1-\beta)\Sigma_2]^{-1}(\mu_2 - \mu_1) + .5 \ln \frac{|\beta\Sigma_1 + (1-\beta)\Sigma_2|}{|\Sigma_1|^\beta |\Sigma_2|^{1-\beta}}$.

b. What will be the Bhattacharya bound in this case? [2] [CO1]