# IT 585 -Advanced Machine Learning

## Quiz 2

March 10, 2023
Duration: 45 minutes
Maximum Marks: 10

***Note: In case you have any doubt, make an appropriate assumption, state the assumption clearly, and proceed. Proofs should be complete***

1. Suppose we have $n$ samples $X_1, X_2, \dots X_n$. $X_i = 1$ if a randomly selected student from DA-IICT campus knows Java Programming and 0 otherwise. Assuming that the $X_i$'s are independent Bernoulli random variables with unknown parameter $p$, derive the maximum likelihood estimator of $p$, the proportion of students on DA-IICT campus who know Java Programming. Now assuming a beta prior for $p$ derive the MAP estimate. for the same. The Bernoulli PMF is given as

$$g(x|p) = p^x(1-p)^{(1-x)}, x \in \{0, 1\}$$

and Beta distribution with parameters $\alpha$ and $\beta$ is given as

$$\text{Beta}(p|\alpha, \beta) = C * p^{\alpha-1}(1-p)^{\beta-1}$$

where C is a constant **[3 + 4 =7 Marks]**

2. Let $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ be the given training examples for a regression problem. Assume that for any given data point $\mathbf{x}_i$, nature samples its corresponding value for $y_i$ independently by means of the following distribution:

$$P_{\text{data}}(y_i \mid \mathbf{x}_i) = \mathcal{N}(y_i \mid \boldsymbol{w}^{*\top}\mathbf{x}_i, 1/\alpha^*)$$

where, $\boldsymbol{w}^*$ and $\alpha^*$ are unknown to us. Here $\frac{1}{\alpha^*}$ is variance. Suppose, we learn the parameters $\boldsymbol{w}^*$ and $\alpha^*$ by means of Maximum Likelihood Estimate given by $\boldsymbol{w}_{\text{MLE}}(\mathcal{D})$ and $\alpha_{\text{MLE}}(\mathcal{D})$. Observe that for any fixed set of training examples $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, the maximum likelihood estimates $\boldsymbol{w}_{\text{MLE}}(\mathcal{D})$ and $\alpha_{\text{MLE}}(\mathcal{D})$ depend on the training data $\mathcal{D} = \{y_1, \dots, y_n\}$.

For this problem, assume that data matrix $\boldsymbol{X}$ is full column rank (i.e. $\boldsymbol{X}^\top\boldsymbol{X}$ is invertible). Show that the following property holds true under this setup.

(a) $\mathrm{V}ar_{\mathcal{D}}\left[\boldsymbol{w}_{\text{MLE}}(\mathcal{D})\right] = \frac{\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}}{\alpha^*}$. **[3 Marks]**

[Hint: Expectation and variance needs to be taken w.r.t. probability distribution on the set $\mathcal{D} = \{y_1, \dots, y_n\}$ which would involve $P_{\text{data}}(\cdot)$].