Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-3
Date: October 18, 2023

Instructions:

Total: **30 marks**

1. Duration is 60 minutes. Please write your name, roll number, department on **all pages**.
2. Write your answers clearly in the provided box. Keep your answer precise and concise.

Section 1 (Short/medium-length answer questions: 30 marks). .

1. Given N coin-toss outcomes $\{y_1, y_2, \ldots, y_N\}$, assuming $p(y_n|\theta) = \text{Bernoulli}(y_n|\theta)$, with $y_n = 1$ denoting a heads outcome, the MLE solution is $\theta_{MLE} = \frac{\sum_{n=1}^{N} y_n}{N}$. Further assuming a $\text{Beta}(\theta|\alpha, \beta)$ prior on θ, the MAP solution is $\theta_{MLE} = \frac{\sum_{n=1}^{N} y_n + \alpha - 1}{N + \alpha + \beta - 2}$. Comparing these results, briefly describe what the prior's hyperparameters α and β signify/denote in this problem? **(2 marks)**

2. Assume a prior $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{w}_0, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} \exp\{-\frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{w} - \boldsymbol{w}_0)\}$ on the weight vector of a linear model (assume prior's hyperparameters $\boldsymbol{w}_0, \boldsymbol{\Sigma}$ to be known). What does this specific prior say about $\boldsymbol{w}$? Also write down the expression of the corresponding regularizer on $\boldsymbol{w}$. **(2 marks)**

3. Is logistic regression a generative model or a discriminative model? Briefly explain. Also answer the same question for probabilistic linear regression with Gaussian likelihood for the responses. **(2 marks)**

4. In generative classification model with Gaussian class-conditionals, what are the parameters that we need to estimate? Also write down the expression, in terms of these parameters, for probability of a test input $\boldsymbol{x}_*$ belonging to class $k \in \{1, 2, \ldots, K\}$ (your answer up to a proportionality constant is fine). **(2 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-3
Date: October 18, 2023

5. Given the learned means $\mu_1, \mu_2, \ldots, \mu_K$ from a K-means clustering model, how would you obtain a soft clustering for input $\boldsymbol{x}_n$. Write your answer in terms of the precise mathematical expression. **(2 marks)**

6. If you learn a Gaussian mixture model (assuming K Gaussians) using the ALT-OPT algorithm, will the expression for the hard guess of the cluster id $\boldsymbol{z}_n$ for an input $\boldsymbol{x}_n$ be the same as what you would have in K-means clustering algorithm? Briefly explain your answer. **(2 marks)**

7. Given a Gaussian posterior distribution $\mathcal{N}(\theta|\mu, \sigma^2)$ for some parameter $\theta \in \mathbb{R}$, what is the MAP solution? Given this Gaussian posterior, can you obtain the MLE solution? Briefly explain your answer. **(2 marks)**

8. Briefly explain the basic idea of multi-dimensional scaling (MDS). Why is MDS not suitable for getting out-of-sample embedding? **(2 marks)**

9. Write down the expression for an appropriate distortion/reconstruction error function for a matrix factorization problem in which we wish to factorize an $N \times M$ matrix $\mathbf{X}$ as a product of an $N \times K$ matrix $\mathbf{Z}$ and a $K \times M$ matrix $\mathbf{W}$. Briefly explain how would you solve for $\mathbf{Z}$ and $\mathbf{W}$? **(2 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-3
Date: October 18, 2023

10. In PCA, what top-most eigenvector (i.e., which corresponds to the largest eigenvalue) means? **(2 marks)**

11. Briefly explain why PCA might not be a suitable dimensionality reduction method if our eventual goal is to learn a classification model using the lower-dimensional inputs obtained by PCA? **(2 marks)**

12. Briefly explain the basic idea of K-means++. **(2 marks)**

13. Briefly explain what is semi-supervised learning and how can we use a latent variable model for semi-supervised learning. **(2 marks)**

14. For a latent variable model with likelihood $p(\boldsymbol{x}_n|\boldsymbol{z}_n, \theta)$, prior $p(\boldsymbol{z}_n|\phi)$ on the latent variables $\boldsymbol{z}_n$, write down the expression the incomplete data log-likelihood (ILL) $p(\boldsymbol{x}_n|\theta, \phi)$, assuming $\boldsymbol{z}_n$ is discrete. **(2 marks)**

15. For a dimensionality reduction method, briefly explain what the encoder and decoder components of the model are used for? **(2 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-3
Date: October 18, 2023