Introduction to ML (CS771), 2024-2025-Sem-I		Total Marks	25
Quiz 3. October 24, 2024		Duration	45 minutes
Name		Roll No.	

Instructions:

1.	Clearly write your name (in block letters) and roll number in the provided boxes above.
2.	Write your final answers concisely in the provided space. You may use blue/black pen.
3.	We won't be able to provide clarifications during the quiz. If any aspect of some question appears ambiguous/unclear to you, please state your assumption(s) and answer accordingly.

Question 1: Write **T** or **F** for True/False in the box next to each question given below, with a brief (1-2 sentences at most) explanation in the provided space in the box below the question. Marks will be awarded only when the answer (T/F) and explanation both are correct. (**3 x 2 = 6 marks**)

1.1	To predict the label using a generative classification model, comparing the probabilities $p(y = k\|\boldsymbol{x})$ for different values of k is equivalent to comparing the class-conditional probability densities $p(\boldsymbol{x}\|y = k)$ for different values of k	

1.2	A Gaussian prior $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}\|\boldsymbol{w}_0, \lambda^{-1}\boldsymbol{I})$ on the weight vector $\boldsymbol{w} \in \mathbb{R}^D$ will cause a regularization effect and encourage the entries in $\boldsymbol{w}$ to take small values.	

1.3	Even though the MAP estimate is the mode of the posterior distribution, to compute the MAP estimate, it is not necessary to compute the posterior distribution.	

Question 2: Answer the following questions concisely in the space provided below the question.

2.1	Consider the RBF kernel $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right)$ where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are D dim inputs. Consider two cases: (1) when bandwidth hyperparameter γ is set as very-very large, and (2) when γ is set as very-very small. For each of these two cases, answer (with brief justification) whether the resulting kernel function would be practically useful. **(4 marks)**

2.2	Briefly explain why using kernels with the landmarks approach or the random features approach is faster at test time than using kernels in the standard manner? **(3 marks)**

2.3	Given a dataset $\boldsymbol{X}$ as the $N \times D$ input matrix with N inputs and D features, write down the K-means hard-clustering problem for this dataset in form of an equivalent matrix factorization problem, clearly specifying the meanings of the variables involved in the matrix factorization, their dimensions, and constraints on them, if any. **(4 marks)**

2.4	Why is it difficult to compute the predictive distribution of a logistic regression model which, by definition, is given by $p(y_* = 1|\boldsymbol{x}_*, \boldsymbol{X}, \boldsymbol{y}) = \int p(y_* = 1|\boldsymbol{w}, \boldsymbol{x}_*)p(\boldsymbol{w}|\boldsymbol{X}, \boldsymbol{y})d\boldsymbol{w}$. Suggest a method to approximate it and clearly show the necessary equations. **(3 marks)**

2.5	Show that, for generative classification with uniform class marginal and Gaussian class conditionals $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, the posterior probability of input $\boldsymbol{x}$ belonging to class k, i.e., $p(y = k|\boldsymbol{x}) \propto \exp(\boldsymbol{w}_k^\top x + b_k)$, and write down the expressions for $\boldsymbol{w}_k$ and b_k **(5 marks)**