Introduction to ML (CS771), 2024-2025-Sem-I		Total Marks	25
Quiz 1. August 20, 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	Learning with Prototypes (LwP) based classification is, in general, faster at test time than *K*-nearest neighbors based classification	

1.2	Large Euclidean distance between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ implies their high similarity.	

1.3	Prediction cost (time taken to predict the label of a test input) for a decision tree (DT) is proportional to the size of the training data that was used to construct the DT.	

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

2.1	Consider two column vectors $\boldsymbol{a} = [a_1, a_2]$ and $\boldsymbol{b} = [b_1, b_2]$, each living in a two-dimensional vector space. Show that a similarity function defined as $k(\boldsymbol{a}, \boldsymbol{b}) = (1 + \boldsymbol{a}^\top \boldsymbol{b})^2$ is equivalent to the standard dot product but in a new vector space with both vectors transformed by a mapping function $\boldsymbol{\phi}$, i.e., $k(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{\phi}(\boldsymbol{a})^\top \boldsymbol{\phi}(\boldsymbol{b})$. Clearly specify the form of the mapping $\boldsymbol{\phi}$. What is the dimensionality of this new vector space? **(4 marks)**

2.2 To speed up distance computations in *K*-nearest neighbors (*K*NN), consider two methods to reduce the dimensionality of the inputs: (1) Selecting 10 features (say using some feature selection method) from the originally given features, and (2) Applying some transformation such that each input is converted into a 10-dimensional binary vector. Which of these two methods would potentially be more suited for faster predictions in *K*NN, and why? **(3 marks)**

2.3 Suppose we want to learn a decision tree (DT) from some toy training data shown on the right side. Each row is a training example has 3 continuous-valued features and a binary label. Show a suitable DT given this training data and briefly justify why your DT would be ideal for this data. **(3 marks)**

F1	F2	F3	Y
1	2	3	-1
2	4	6	-1
3	6	9	+1
4	8	12	+1

2.4 If prediction speed is not an issue, can internal nodes of a DT split data based on ALL features but not using criteria such as information gain? Brief explain your answer. **(3 marks)**

2.5 What effect would minimizing a regularized loss function $L(\boldsymbol{w}) + \lambda R(\boldsymbol{w})$ have on the weight vector $\boldsymbol{w} \in \mathbb{R}^D$ if $R(\boldsymbol{w})$ is an ℓ_2 squared norm, i.e., $R(\boldsymbol{w}) = \boldsymbol{w}^\top \boldsymbol{w}$, and why? **(3 marks)**

2.6 Suppose you have trained two classifiers- LwP and DT - on the same training data with 5 classes. In future, suppose training data from 3 more classes becomes available (so now we have 8 classes). Which of these two models will be easier to retrain, and why? **(3 marks)**