INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

Mid-Term Examination (MTE)

Machine Learning (CSN-526)

Spring Semester 2021-22

Time: 60 minutes **Total Marks: 100**

Part: 1

True/False Questions: Questions 01 – 20 (20x3=60 marks).

Question 1:
Classifying natural numbers into prime vs non-prime is a good example of a problem that can be solved using machine learning (e.g. by binary classification)

Question 2:

Learning with prototypes cannot be used if there are more than two classes

Question 3:
If $f, g: \mathbb{R} \to \mathbb{R}$ are two convex functions (not necessarily differentiable), then the average function $h(x) \triangleq \left(f(x) + g(x)\right)/2$ must always be convex as well

Question 4:
Let $f: \mathbb{R} \to \mathbb{R}$ be a doubly differentiable function (i.e. first and second derivatives exist). If $f'(x^0) = 0$ at $x^0 \in \mathbb{R}$, then it must always be true that $f''(x^0) = 0$

Question 5:

The boundary of the unit 2D circle i.e. $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is a convex set

Question 6:
When minimizing a convex function f using (sub)gradient descent (without any constraints), it does not matter what step lengths we choose since f is convex

Question 7:

Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined as $f(x) = 2x^2 + x + 2$, the subdifferential of f at $x = -1$ i.e. $\partial f(-1)$ is the set $\{-3, 3\}$

Question 8:

Executing one step of mini-batch stochastic gradient descent with large batch size is usually more expensive that executing one step of stochastic gradient descent

Question 9:

The Hessian of a function $f: \mathbb{R}^2 \to \mathbb{R}$ is always a 2×2 PSD matrix

Question 10:

If $f, g: \mathbb{R} \to \mathbb{R}$ are two non-differentiable functions, then the function $f + g$ will always be non-differentiable as well

Question 11:

Suppose X is a random variable such that $x \geq 1$ for all $x \in S_X$ where S_X is the support of X. Then the variance of X must be greater than 1 too i.e. $\mathbb{V}[X] \geq 1$

Question 12:

The Adagrad method is a technique for choosing an appropriate batch size when training a deep network.

Question 13:

When using kNN to do classification, using a large value of k always gives better performance since more training points are used to decide label of the test point

Question 14:

Cross validation means taking a small subset of the test data and using it to get an estimate of how well will our algorithm perform on the entire test dataset

Question 15:

If X and Y are two real-valued random variables such that $\mathrm{Cov}(X, Y) < 0$ then at least one of X or Y must have negative variance i.e. either $\mathbb{V}X < 0$ or $\mathbb{V}Y < 0$

Question 16:

Suppose X is a real valued random variable with variance $\mathbb{V}X = 9$. Then the random variable Y defined as $Y = X - 2$ will always satisfy $\mathbb{V}Y = \mathbb{V}X - 2^2 = 5$

Question 17:

The LwP algorithm for binary classification always gives linear decision boundary if we use one prototype per class and Euclidean distance to measure distances

Question 18:

If $f, g: \mathbb{R}^2 \to \mathbb{R}$ are two non-convex functions, then the function $h: \mathbb{R}^2 \to \mathbb{R}$ defined as $h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ must always be non-convex too

Question 19:

The LwP algorithm when used on a binary classification problem, results in a linear decision boundary no matter how many prototypes we use per class

Question 20:

The time it takes to make a prediction for a test data point with a decision tree with n leaf nodes is always $\mathcal{O}(\log n)$ no matter what the structure of the tree.

Part: 2

Choose all the correct options (many may be correct): Questions 21 – 28 (8x5=40 marks).

Question 21:

Suppose $f, g: \mathbb{R} \to \mathbb{R}$ are two convex functions (not necessarily differentiable) and we define $p(x) = f(x) + g(x)$ and $q(x) = f(x) - g(x)$. Two claims are made about these functions

Claim 1: $p(x) + q(x)$ must always be convex

Claim 2: $q(x)$ must always be convex

A	Claim 1 is TRUE, claim 2 is FALSE
B	Claim 1 is FALSE, claim 2 is TRUE
C	Both claims are TRUE
D	Both claims are FALSE

Question 22:

Let $f(x) = \sin(x)$. Which of the following statements is true about the function $f(x)$?

A	$f(x)$ has more than one local minima
B	$f''''(x) = f(x)$
C	$f(x)$ is a convex function
D	$f(x)$ is a concave function

Question 23:

Which of the following statements is true about the kNN algorithm?

A	When used for binary classification, the kNN algorithm always produces decision boundaries that are linear (i.e. a line or a hyperplane)
B	The kNN algorithm can be used to solve regression problems
C	The value k in kNN must always be a positive integer
D	There exists no dataset, nor any value of k, for which kNN has linear decision boundary

Question 24:

Which of the following statements is true?

A	In held-out validation, the validation set is a subset of the test set
B	In held-out validation, the validation set is a subset of the training set
C	Using Multi fold Cross Validation is more expensive than held-out validation
D	Using Multi fold Cross Validation is less expensive than held-out validation

Question 25:

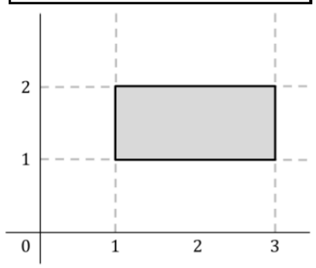
Let X, Y be two r.v. (not necessarily independent) with same support $S_X = \{1,2\} = S_Y$. Let $P \in \mathbb{R}^{2\times 2}$ with $P_{ij} = \mathbb{P}[X = i, Y = j], i, j \in \{1,2\}$ encode the joint PMF where P_{ij} is the element in the i^{th} row and j^{th} column. Let $\mathbf{1} = [1,1] \in \mathbb{R}^2$ denote the all ones vector. Which of the following is true?

A	$P\mathbf{1}$ gives marginal PMF of X
B	$P^\top\mathbf{1}$ gives marginal PMF of X
C	$P\mathbf{1}$ gives marginal PMF of Y
D	$P^\top\mathbf{1}$ gives marginal PMF of Y

Question 26:

Consider the following optimization problem.

$$\max_{(x,y)\in\mathbb{R}^2} x^2 + y^2$$
$$x \in [1,3]$$
$$y \in [1,2]$$



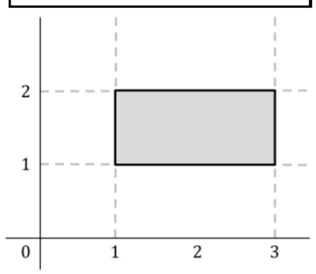
At which point in $\mathbb{R}^2$ is the solution to this optimization problem achieved?

A	**(1,1)**
B	**(3,1)**
C	**(1,2)**
D	**(3,2)**

Question 27:

Consider the following optimization problem.

$$\max_{(x,y)\in\mathbb{R}^2} -x^2 - y^2$$
$$x \in [1,3]$$
$$y \in [1,2]$$



At which point in $\mathbb{R}^2$ is the solution to this optimization problem achieved?

A	**(1,1)**
B	**(3,1)**
C	**(1,2)**
D	**(3,2)**

Question 28:

Ridge regression uses what penalty on the regression weights?

(a) L_0
(b) L_1
(c) L_2
(d) L_2^2