Questions List With Marks

<< Go back to the results page (https://hello.iitk.ac.in/manage/cs771a/node/36/results?page=0)

Starting Time: October 5, 2020 18:16 **Submitted Time: October 5, 2020 18:44**

Question:

L1 norm distance between two vectors a = [3,5,1] and b = [5,2,5] is

Options and Responses:

- ✖ -3
- ✖ 6
- ✔ 9 (selected by student)
- ✖ 5

Score: 1

Question:

Learning with prototypes (LwP) can be used for following problems (check all that apply)

Options and Responses:

- ✖ Dimensionality reduction
- ✔ Binary classification (selected by student)
- ✔ Multi-class classification (selected by student)
- ✖ Regression (selected by student)

Score: 0

Question:

LwP with Mahalanobis distance can learn nonlinear decision boundaries

Correct Answer and Responses:

Correct Answer: False

Answer Given: False

Score: 1

Question:

Assuming binary classification problem with N training examples (assuming we have examples from both classes) and using Euclidean distance, LwP will learn the same decision boundary as KNN if (check all that apply)

Options and Responses:

- ✖ N is very large
- ✔ K=1
- ✔ N=2
- ✖ Such a thing will never happen (selected by student)

Score: 0

Question:

A decision tree can be used for

Options and Responses:

- ✖ Clustering (selected by student)
- ✔ Classification (selected by student)
- ✔ Regression (selected by student)
- ✖ Dimensionality reduction (selected by student)

Score: 0

Question:

For regression with decision trees, information gain can't be used as a splitting criterion but gini index can be used

Correct Answer and Responses:

Correct Answer: False

Answer Given: False

Score: 1

Question:

Decision trees cannot be used with real-valued features

Correct Answer and Responses:

Correct Answer: False

Answer Given: False

Score: 1

Question:

For a linear regression model (ignoring the bias term) with N training examples having D features each, the model size will be (in terms of the total number of scalars)

Options and Responses:

- ✖ N
- ✖ N*D
- ✔ D (selected by student)
- ✖ constant (independent of N and D)

Score: 1

Question:

Which of the following regression loss functions are differentiable everywhere?

Options and Responses:

- ✖ epsilon-insensitive loss
- ✖ absolute loss
- ✔ squared loss (selected by student)
- ✖ Huber loss (selected by student)

Score: 0

Question:

Which of these regularization methods promote sparse solutions?

Options and Responses:

- ✔ L1 (selected by student)
- ✖ L2
- ✔ L0
- ✖ early stopping

Score: 0

Question:

Which of these are convex functions?

Options and Responses:

- ✔ 2*x + 3 (selected by student)
- ✔ -2*x + 3 (selected by student)
- ✔ x*x (selected by student)
- ✖ x*x*x

Score: 1

Question:

Which of these statements about gradient descent (GD) are true?

Options and Responses:

✖ If run for sufficiently long, it is guaranteed to find the global minima

✔ It is sensitive to initialization (selected by student)

✔ Every step of GD moves in the opposite direction to the current gradient.

(selected by student)

✖

When optimizing vector-valued variables, it optimizes one element of the vector at a time.

Score: 1

Question:

Which of the following is true about the absolute value function f(x) = |x|

Options and Responses:

✔ It has infinite many subgradients in its subdifferential set at x = 0

(selected by student)

✖

It has a very large but finite number of subgradients in its subdifferential set at point x = 0

✔ It is differentiable everywhere except x = 0 (selected by student)

✔ It is a convex function (selected by student)

Score: 1

Question:

At test time, a decision tree with a single decision node (with a single feature's value based or an LwP based splitting criterion) will be faster than a one-nearest neighbors method

Correct Answer and Responses:

Correct Answer: True

Answer Given: True

Score: 1

Question:

For which of these models, the test time cost (time it takes to make a prediction for a test example) will increase if we increase the training set size?

Options and Responses:

- ✖ LwP
- ✔ Nearest neighbors (selected by student)
- ✖ Decision tree (assuming a constant prediction rule at the leaf nodes)
- ✖ Ridge regression

Score: 1

Question:

For unconstrained problems, gradient descent and projected/proximal gradient descent will give the same solution.

Correct Answer and Responses:
Correct Answer: True

Answer Given: True

Score: 1

Question:

For convex functions, Newtons method has the same per-iteration time-cost as gradient descent.

Correct Answer and Responses:
Correct Answer: False

Answer Given: False

Score: 1

Question:

Increasing the extent of regularization (e.g., the value of regularization hyperparameter) may not necessarily increase the validation set accuracy

Correct Answer and Responses:

Correct Answer: True

Answer Given: True

Score: 1

Question:

A linear regression model with L1 norm regularizer will not have a closed form expression for the optimal weight vector.

Correct Answer and Responses:

Correct Answer: True

Answer Given: True

Score: 1

Question:

Assuming binary classification and each input to be 10 dimensional, the **minimum** number of parameters (in terms of the number of scalar values) to store an LwP model will be:

(Note: Do not assume that the input features have been transformed/augmented (each will be 10 dimensional)_

Options and Responses:

- ✖ 10
- ✖ 20 (selected by student)
- ✖ 22
- ✔ 11

Score: 0