0/35 Questions Answered

Homework 2

Q1 Regression
8 Points

Answer multiple-choice questions on regression. Note that there are no resubmissions, so be sure to print out, answer by hand, and only enter the response when ready for grading.

Q1.1
1 Point

True-False: Linear Regression is a supervised machine learning algorithm.

True

False

Save Answer

Q1.2
1 Point

True-False: Linear Regression is mainly used for Regression.

True

False

Save Answer

Q1.3
1 Point

Which methods do we use to find the best-fit line for data in Linear Regression?

Least Square Error

Maximum Likelihood

Logarithmic Loss

Both A and B

Save Answer

Q1.4
1 Point

Which evaluation metrics can evaluate a model while modeling a continuous output variable?

AUC-ROC

Accuracy

Logloss

Mean-Squared-Error

Save Answer

Q1.5
1 Point

Which of the following statement is true about outliers in Linear regression?

Linear regression is sensitive to outliers

Linear regression is not sensitive to outliers

Can’t say

None of these

Save Answer

Q1.6
1 Point

Suppose that we have N independent variables (X1, X2, ..., Xn) and the dependent variable is Y. Now, Imagine that you are applying linear regression by fitting the best-fit line using the least square error on this data. You found that the correlation

coefficient for one of its variables (Say X1) with Y is -0.95. Which of the following is valid for X1?

The relation between the X1 and Y is weak

The relation between the X1 and Y is strong

The relation between the X1 and Y is neutral

Correlation can't judge the relationship

Save Answer

Q1.7
1 Point

We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation." Which of the following is/are true about Normal Equations?
(1) We don't have to choose the learning rate.
(2) It becomes slow when the number of features is very large.
(3) There is no need to iterate.

1 and 2

1 and 3

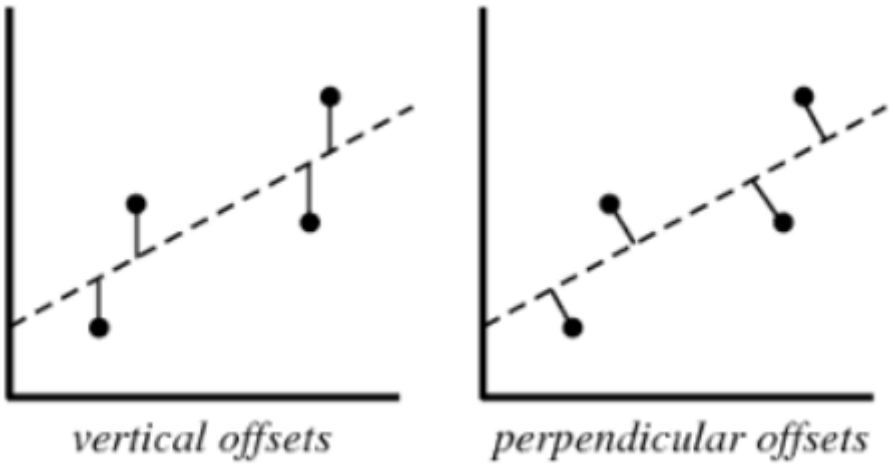
2 and 3

1,2 and 3

Save Answer

Q1.8
1 Point

Which offsets do we use in linear regression's least square line fit? Suppose the horizontal axis is the independent variable and the vertical axis is the dependent variable.



Vertical offset

Perpendicular offset

Both, depending on the situation

None of above

Save Answer

Q2 Polynomial Regression
2 Points

Answer multiple-choice questions. Note that there are no resubmissions, so be sure to print out, answer by hand, and only enter the response when ready for grading.

Q2.1
1 Point

In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?

Bias will be high, and variance will be high

Bias will be low, and variance will be high

Bias will be high, and variance will be low

Bias will be low, and variance will be low

Save Answer

Q2.2
1 Point

What will happen when you fit the degree 4 polynomial in linear regression?

There is a high chance that the degree 4 polynomial will overfit the data

There is a high chance that degree 4 polynomial will underfit the data

Can’t say

None of these

Save Answer

Q3 Use-Case
2 Points

Suppose you have a dataset D1 and design a linear regression model of degree 3 polynomial. You find that the training and testing error is “0” or, in another term, it perfectly fits the data.

Q3.1
1 Point

What will happen when you fit the degree 4 polynomial in linear regression?

There is a high chance that the degree 4 polynomial will overfit the data

There is a high chance that degree 4 polynomial will underfit the data

Can’t say

None of these

Save Answer

Q3.2
1 Point

What will happen when you fit the degree 2 polynomial in linear regression?

It is a high chance that degree 2 polynomial will overfit the data

It is a high chance that degree 2 polynomial will underfit the data

Can’t say

None of these

Save Answer

Q4 Other Regression Types
4 Points

Q4.1
1 Point

True-False: Lasso Regularization can be used for variable selection in Linear Regression.

True

False

Save Answer

Q4.2
1 Point

What will happen when you apply a very large penalty in the case of Lasso?

Some of the coefficients will become zero

Some of the coefficients will be approaching zero but not absolute zero

Both A and B, depending on the situation

None of these

Save Answer

Q4.3
1 Point

What will happen when you apply a very large penalty in Ridge Regression?

Some of the coefficients will become absolute zero

Some of the coefficients will approach zero but not absolute zero

Both A and B depending on the situation

None of these

Save Answer

Q4.4
1 Point

Choose the option which describes bias in the best manner for Ridge Regression.

In the case of a very large x, bias is low

In the case of a very large x, bias is high

We can't say about bias

None of these

Save Answer

Q5 Bias vs Variance
5 Points

Q5.1
1 Point

True- False: Overfitting is more likely when you have a lot of data to train.

TRUE

FALSE

Save Answer

Q5.2
1 Point

True-False: Overfitting occurs when a model is too complex for the given data, and its performance on the training data is much better than on the test data.

TRUE

FALSE

Save Answer

Q5.3
1 Point

True-False: Regularization is a technique used to prevent overfitting in machine learning models.

TRUE

FALSE

Save Answer

Q5.4
1 Point

True-False: The bias-variance tradeoff is between a model's ability to fit the data well and generalize to new, unseen data.

TRUE

FALSE

Save Answer

Q5.5
1 Point

True-False: The ROC curve is a commonly used evaluation metric for regression problems.

TRUE

FALSE

Save Answer

Q6 Regression in Practice
5 Points

We have been given a dataset with n records in which we have an input attribute as x and an output attribute as y. Suppose we use a linear regression method to model this data. We randomly split the data into training and test sets to test our linear regressor.

Q6.1
1 Point

Now we increase the training set size gradually. What will happen with the mean training error as the training set size increases?

Increase

Decrease

Remain constant

Can’t Say

Save Answer

Q6.2
1 Point

What do you expect will happen with bias and variance as you increase the size of training data?

Bias increases, and Variance increases

Bias decreases, and Variance increases

Bias decreases, and Variance decreases

Bias increases and Variance decreases

Can’t Say False

Save Answer

Q6.3
1 Point

True-False: A validation dataset is used to evaluate a final tuned model.

True

False

Save Answer

Q6.4
1 Point

True-False: The test dataset is used for hyperparameter tuning.

True

False

Save Answer

Q6.5
1 Point

True-False: The notion of a validation set and test dataset may disappear with k-fold cross-validation.

True

False

Save Answer

Q7 Experiments and Data
18 Points

Short answers. Provide 1-2 sentences in response to each prompt.

Be sure to read the PDF provided as part of this assignment:
https://canvas.tufts.edu/courses/44718/files/5646191?module_item_id=982771

Q7.1
2 Points

What is the purpose of model evaluation?

Enter your answer here

Save Answer

Q7.2
2 Points

What is the drawback of training and testing on the same data?

Enter your answer here

Save Answer

Q7.3
2 Points

How does train/test split work, and what is its primary drawback?

Enter your answer here

Save Answer

Q7.4
2 Points

How does K-fold cross-validation work, and what is the role of ”K”?

Enter your answer here

Save Answer

Q7.5
2 Points

Why do we pass X and y, not X train and y train, to the ‘cross val score‘ function?

Enter your answer here

Save Answer

Q7.6
2 Points

Why does a ‘cross val score‘ need a ”scoring” parameter?

Enter your answer here

Save Answer

Q7.7
2 Points

What does ‘cross val score‘ return, and what do we usually do with that object?

Enter your answer here

Save Answer

Q7.8
2 Points

Under what circumstances does ‘cross val score‘ return negative scores?

Enter your answer here

Save Answer

Q7.9
2 Points

When should you use train/test split, and when should you use cross-validation?

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission ❯