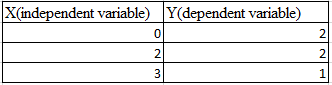
**Q1. Which of the following step / assumption in regression modeling impacts the trade-off between under-fitting and over-fitting the most.**

A. The polynomial degree B. Whether we learn the weights by matrix inversion or gradient descent C. The use of a constant-term

**Q2. Suppose you have the following data with one real-value input variable & one real-value output variable. What is leave-one out cross validation mean square error in case of linear regression (Y = bX+c)?**



A. 10/27 B. 20/27 C. 50/27 D. 49/27

Q3. Which of the following is/ are true about “Maximum Likelihood estimate (MLE)”?

1. MLE may not always exist
2. MLE always exists
3. If MLE exist, it (they) may not be unique
4. If MLE exist, it (they) must be unique

A. 1 and 4 B. 2 and 3 C. 1 and 3

Q4. Which of the following indicates a fairly strong relationship between X and Y?

A. Correlation coefficient = 0.9

B. The p-value for the null hypothesis Beta coefficient =0 is 0.0001

C. The t-statistic for the null hypothesis Beta coefficient=0 is 30

D. None of these

Q5. Suppose you are training a linear regression model. Now consider these points.

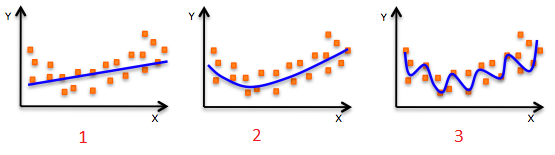
1. Overfitting is more likely if we have less data
2. Overfitting is more likely when the hypothesis space is small

Which of the above statement(s) are correct?

A. Both are False B. 1 is False and 2 is True C. 1 is True and 2 is False

D. Both are True

Q6. The following visualization shows the fit of three different models (in black line) on same training data. What can you conclude from these visualizations?



1. The training error in first model is higher when compared to second and third model.
2. The best model for this regression problem is the last (third) model, because it has minimum training error.
3. The second model is more robust than first and third because it will perform better on unseen data.
4. The third model is overfitting data as compared to first and second model.
5. All models will perform same because we have not seen the test data.

A. 1 and 3 B. 1 and 3 C. 1, 3 and 4 D. Only 5

Q7. The expected value of Y is a linear function of the X(X1,X2….Xn) variables and regression line is defined as:

Y = β0 +  β1 X1 + β2 X2……+ βn Xn

Which of the following statement(s) are true?

1. If Xi changes by an amount ∆Xi, holding other variables constant, then the expected value of Y changes by a proportional amount βi ∆Xi, for some constant βi (which in general could be a positive or negative number).
2. The value of βi is always the same, regardless of values of the other X’s.
3. The total effect of the X’s on the expected value of Y is the sum of their separate effects.

Note: Features are independent of each others(zero interaction).

A. 1 and 2 B. 1 and 3 C. 2 and 3 D. 1,2 and 3

Q8. How does number of observations influence overfitting? Choose the correct answer(s). Note: Rest all parameters are same

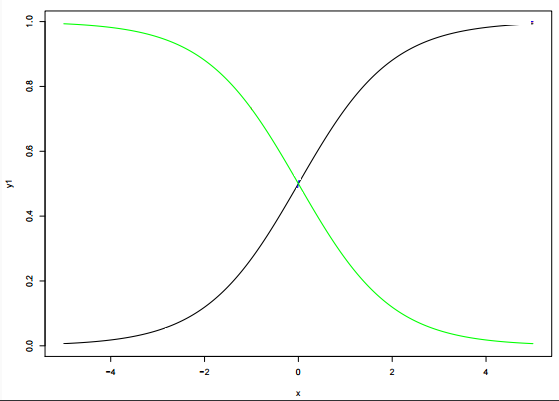
1. In case of fewer observations, it is easy to overfit the data.
2. In case of fewer observations, it is hard to overfit the data.
3. In case of more observations, it is easy to overfit the data.
4. In case of more observations, it is hard to overfit the data.

A. 1 and 4 B. 2 and 3 C. 1 and 3 D. None of theses

Q9. Logistic Regression transforms the output probability to be in a range of [0, 1]. Which of the following function is used by logistic regression to convert the probability in the range between [0,1].

A. Sigmoid B. Mode C. Square D. Probit

Q10. Below are two different logistic models with different values for β0 and β1.



Which of the following statement(s) is true about β0 and β1 values of two logistics models (Green, Black)?

Note: consider Y = β0 + β1\*X.  Here, β0 is intercept and  β1 is  coefficient.

A. β1 for Green is greater than Black B. β1 for Green is lower than Black

C. β1 for both models is same D. Can’t Say.

Q#11 True or False: A fitted model with more predictors will necessarily have a lower Training Set Error than a model with fewer predictors.

Q#12 While doing a homework assignment, you fit a Linear Model to your data set. You are thinking about changing the Linear Model to a Quadratic one. Which of the following is most likely true:

a) Using the Quadratic Model will decrease your Irreducible Error.

b) Using the Quadratic Model will decrease the Bias of your model. Using the Quadratic Model will decrease the Bias of your model.

c) Using the Quadratic Model will decrease the Variance of your model

d) Using the Quadratic Model will decrease your Reducible Error

Q#13 For each of the following parts, indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible model.

1. The sample size n is extremely large, and the number of predictors p is small:
2. Flexible is worse b) Flexible is better
3. The number of predictors p is extremely large, and the sample size n is small:

A) Flexible is worse b) Flexible is better

iii) The relationship between the predictors and response is highly non-linear:

1. Flexible is worse b) Flexible is better

vi) The variance of the error terms, i.e. σ2=Var(ϵ), is extremely high:

1. Flexible is worse b) Flexible is better

Q#14 Which of the following tools would be well suited for predicting if a student will get an A in a class based on the student's height, and parents’ income? Select all that apply:

1. Linear Discriminant Analysis b) Linear Regression c) Logistic Regression d) Random Guess

Q#15 If we use ten-fold cross-validation as a means of model selection, the cross-validation estimate of test error is:

1. biased upward b) biased downward c) unbiased e) potentially any of the above

Q#2

1. Explain LOOCV and k-fold cross validation. How do these are implemented.

b) What are the advantage and disadvantage of k-fold cross-validation relative to

1. The Validation set approach? 2. LOOCV?

c) Carefully explain the differences between the KNN classifier and KNN regression methods.(5 Pts)

Question#3 Examine the differences between LDA and QDA. (10 Points)

1. If the Bayes decision boundary is linear, do we expect LDA or QDA to perform better on the training set? On the test set?
2. If the Bayes decision boundary is non-linear, do we expect LDA or QDA to perform better on the training set? On the test set?
3. In general, as the sample size n increases, do we expect the test prediction accuracy of QDA relative to LDA to improve, decline, or be unchanged? Why?

Best of Luck