

# Q53-WilliamKennedy-300015367

William Kennedy

2023-12-10

1. How does the bias-variance decomposition of a ridge regression estimator compare with that of ordinary least squares regression?

- i) False, it reduces the variance
- ii) True, since it reduces the variance it will have a higher bias.
- iii) False,
- iv) False

ChatGPT contradicts itself by saying both ii) and iii) are true

2. Both PCA and LASSO can be used for feature selection and/or dimension reduction. Which of the following statements are true?

- i) True, LASSO will choose  $\max\{p, N\}$  so yes, either a subset of the features or the full feature set is chosen.
- ii) True, the first PCA is the normalized linear combination of the variables with the largest variance.
- iii) False, LASSO deduces how many parameters should be chosen. You do not get to choose. In PCA we can choose the number of principal components but its not obvious what is the best value.
- iv) False, regularizes the data and is a only used in regression not classification.

Hence, ChatGPT got this one correct.

3. Why would we use a random forest instead of a decision tree?

- i) False, when the variance is high the training error is low and the testing error is high.
- ii) True, the main goal of a random forrest is reduce the variance by averaging the bagged predictions.
- iii) False, the aim of a random forrest is to reduce the variance. Hence if the variance is high then the bias low since they are inversely related.
- iv) In a sense, this is true. It's not explicitly true, but bringing down the variance reduces the overall complexity of the model which you can argue makes it easier for a human to understand.

Hence, ChatGPT answered correctly. But it could've argued that it makes it for humans to understand since it reduces the complexity by reducing the variance.

4. The optimal Bayes decision rule with the indicator function:

i)False, this is the decision made Naive Bayes. ii) True, this is the theoretical ideal of the optimal classifier. But is not usually attainable due to the computational complexity being NP-Hard iii) Mostly true, it is the best a classifier can achieve locally iv) False, as said in (ii) from a large sample the computational complexity becomes too difficult

ChatGPT's responses were mostly correct. Except it said i) is true and it's not. Its only true for Naive Bayes

5. Consider a probability-based binary classifier. Which of the following statement(s) is/are always true about the ROC curve, and the area under the ROC curve (AUC):

- i) False
- ii) True
- iii) True, this how it is described in the textbook.
- iv) False

6. Which of the following algorithms can learn nonlinear decision boundaries?

- i) True, the decision boundaries are quadratic functions.
- ii) True, since the division of the feature space is usually not linear.
- iii) True, since logistic regression is considered a special case of polynomial regression.
- iv) False,

ChatGPT got all of these correct.

7. Which of the following are true of binary classification/regression trees?

- i) True, decision trees are highly sensitive to the underlying data.
- ii) True, the intention of a random forest is to reduce the variance and thus reduce overfitting.
- iii) False, bagging does the opposite. It is likely to reduce variance.
- iv) True, a deep decision tree will have more classification classes and specify the observations more strictly.

ChatGPT was mostly correct, however it did not confirm that decision trees are sensitive to changes in data.

8. A regression tree has substantially higher validation MSE than expected. Which of the following is likely to improve validation MSE in most real-world applications?

i) False, adding more features will increase complexity and make it harder to predict new results. ii) True, if the each feature is representative of a group of highly correlated features then this could reduce the overall complexity and the model could have a looser fit. iii) True, tree pruning will reduce complexity and CV would be a good method to use. iv) False,

ChatGPT missed the fact that use a smaller subset of features would reduce complexity and possibly improve the test MSE.

9. A decision tree is getting abnormally bad performance on both the training and test sets. What could be causing the problem?

- i) Possibly true, if the tree is shallow it will have a high bias, hence both the training and test error will be high.
- ii) False, this would tend towards a model that is overfitting. Not underfitting.
- iii) False, if the model was overfit then the training error would be low and the test error would be high.
- iv) False

ChatGPT was incorrect in it's predictions, iii) is not true. This would be underfitting, not overfitting.

10. A dataset has 3 pts:  $A = (0, 2)$ ,  $B = (0, 1)$ ,  $C = (1, 0)$ . The 2-means clustering algorithm is initialized with centers at A and B. Where will the centers converge?

- i) False
- ii) True
- iii) False
- iv) False

11. Consider  $T1$ , a decision stump (a tree with one layer below the root) and  $T2$ , a decision tree that is grown till a maximum depth of 4 (so no more than 3 layers below the root). Which of the following is/are correct?

- i) False
- ii) True
- iii) True
- iv) False

$T1$  is an underfit model so it has a high bias and low variance, whereas  $T2$  is overfit so it has a low bias and high variance. ChatGPT made the correct predictions.

12. Which of the following are true about subset selection?

i) True, it reduces the complexity of the model. ii) True, this is why there are computationally simpler variants forward selection and backward selection iii) True, it's mainly necessary when the number of features is greater than the number of observations. But we do want to find the best subset as all the predictors may not be necessary but sometimes they are. iv) False, ridge regression just makes feature coefficients very small but does not remove them from the analysis.

ChatGPT was mostly correct, I think iii) can be open to more interpretation but it is true that it is generally not required.