

## FE590\_Q2

---

1. One of the benefits of the bootstrap method for cross-validation is that it allows the user to emulate the process of obtaining new sample sets
  - a. True
  - b. False
2. In R, given the data set 'data', the expression "glm(y~., data, family = binomial)"
  - a. Produces a logistic regression of the qualitative variable y onto all other variables in data
  - b. Performs a logistic regression of the qualitative variable y onto the first variable data
  - c. Produces a generalized linear regression of the variable y onto all other variables in data, where all the other variables are specified to have binomial distributions
  - d. Produces a linear prediction of the variable y in data onto all other variables in data
3. Forward subset selection and backward subset selection
  - a. On average, produces the same subsets as best subset selection
  - b. Always produce the same subsets
  - c. May produce different subsets
  - d. Always produce difference subsets
4. Using principal components analysis for the purpose of dimension reduction, say from two dimensions to one dimension. All of the following statements are true EXCEPT

- a. Any point in the original two-dimensional space is represented precisely by a two dimensional vector in the new space
  - b. Generally, the procedure involves a positioning of the new axes such that the data is centered at the origin of the new space
  - c. Any point in the original space is represented by a length along the new axis, plus noise component
  - d. The procedure involves rotating the axes for basis for the data, such that one new axis lies along the direction in which the observations vary most
5. When choosing the best model order, the following criteria always produce the same results
  - a. *Cp* and *BIC*
  - b. *Cp* and *AIC*
  - c. *BIS* and adjusted *R-squared*
  - d. *AIC* and *BIC*
6. For regression trees, the predictor space is divided into *J* distinct, non-overlapping, high-dimensional boxes, in which
  - a. The prediction of the response in box *j* is a majority vote of the response classes of observations that fall into box *j*
  - b. The prediction of the response in box *j* is the mean of the responses of observations that fall into box *j*
  - c. The boxes are chosen so as to distribute the observations as evenly between boxes as possible
  - d. Recursive binary splitting is used to consider every possible partition of the feature space into *J* boxes to find the partition that minimizes the *RSS*
7. In the bootstrap method for cross-validation. Sampling from the available data set is done without replacement
  - a. True

**b. False**

8. Leave-one-out cross-validation (LOOCV)

a. Produces a different estimate of MSE each time it is run, depending on the random seed chosen

**b. Typically has less bias than the validation set approach**

c. Does not require the model to be fit(ted) as many times as the validation set approach

d. Uses approximately half of the available data to fit the model

9. In ridge regression. The penalty takes the form of

**a. An L-2 norm**

10. In k-fold cross-validation

**a. The available set of observations is divided into k groups**

b. The procedure is identical to LOOCV when  $k=1$

c. The available set of observations is divided into groups containing k samples each

d. The procedure generally requires the model to be fit(ted) as many times as LOOCV

11. The lasso regression, all of the following statements are true, EXCEPT

a. The lasso shrinks the coefficient estimates towards zero

**b. The lasso cannot produce sparse models**

c. The lasso can force some of the coefficient estimates to be exactly equal to zero when the tuning parameter is sufficiently large

d. The lasso can be used for automatic variable selection

12. The bootstrap method is able to provide an estimate of standard error of an estimate when one has only a small number of observations because it uses resampling without replacement

- a. True
  - b. False**
13. Generalized additive models are called additive because they calculate a separate function  $f_j$  for each predictor  $x_j$
- a. True**
  - b. False
14. The process of shrinking coefficients of a regression towards zero has the benefit of
- a. Decreasing the bias of the estimates
  - b. Decreasing the computational complexity of the estimation procedure
  - c. Decreasing the variance of the estimates**
  - d. Simultaneously decreasing the variance and bias of the estimates
15. For ridge regression, the data should be standardized because
- a. Coefficient estimates are homogeneous of degree 1, meaning that multiplying a predictor  $X$  by a constant  $c$  leads to a scaling of the coefficient estimates by a factor of  $c$  also**
  - b. Coefficient estimates can change when multiplying a given predictor  $X$  by a constant  $c$
16. Important classes of linear model selection include all of the following EXCEPT
- a. Linear regression**
  - b. Dimension reduction
  - c. Subset selection
  - d. Shrinkage
17. In lasso regression, the penalty takes the form of
- a. An L-1 norm**

18. Cubic regression splines
- a. Include constraints that, for splines on either side of any knot, the values first derivatives and second derivatives of the spline functions are all continuous
19. All of the following are examples of bootstrap techniques we discussed in class EXCEPT
- a. Burrowing
20. Maximal Margin classifiers and support vector classifiers are just special cases of support vector machines
- a. True
  - b. False
21. In ridge regression, the shrinkage parameter (or tuning parameter)
- a. Allows the coefficients to be reduced towards zero
  - b. Allows all coefficients (including the intercept) to be reduced towards zero
  - c. Is used with cross-validation to select the model with minimum bias
  - d. Allows the coefficients of unimportant variables to be set exactly to zero
22. Cross-validation is used to assess training error rate
- a. True
  - b. False
23. Cross-validation can be used to assess the accuracy of models for qualitative variables
- a. True
  - b. False

24. For a model with  $n$  observations and  $p$  predictors. Best subset selection requires
- a.  $2^p$  model tests
  - b.  $N^p$  model tests
  - c.  $Np$  model tests
  - d.  $P^2$  model tests
25. The following are all benefits of generalized additive models (GAMs), EXCEPT
- a. One can examine the effect of each predictor on the response individually while holding all of the other predictors fixed
  - b. GAMs can model non-linear relationship[s that standard linear regression will miss
  - c. GAMs are less computationally demanding than linear regression
  - d. GAMs can potentially make more accurate predictions of the response than linear regression can
26. In R, the statement “`l <- sample(10,5,replace = F)`” produces
- a. 5 random integers in the range from 1 to 10
27. For column vectors  $x$  and  $y$ , all of the following expressions are equivalent for regressing  $y$  onto  $x$  and  $x$ -squared. EXCEPT
- a. `Mod <- lm(y ~ x + x^2)`
  - b. Other answers
28. The validation set approach to cross-validation has all of the following potential drawbacks, EXCEPT
- a. The validation estimate of the test error rate can be highly variable
  - b. The approach is difficult to implement

- c. Only a subset of the observations --- those that are included in the training set rather than in the validation set --- are used to fit the model
  - d. The validation set error rate may tend to overestimate the test error rate for the model fit(ted) on the entire data set
29. Neural networks greatest strength is their interpretability
- a. True
  - b. False
30. In the validation set approach to cross-validation, the available set of observations is divided randomly into two parts, a training set and a validation set. In this case, all of the following are true EXCEPT
- a. The fitted model is used to predict the responses for the observations in the validation set
  - b. The training error rate typically overestimates the test error rate
  - c. The model is fit(ted) on the training set
  - d. The validation set error rate provides an estimate of the test error rate
31. Supposed that “wage”...
- a.  $\text{Wage} = \beta_0 + f(\text{year}) + g(\text{age}) + h(\text{education}) + \epsilon$