ORIE 4740 Spring 2017

Sample Exam Questions

Instruction: Check the boxes next to the right answers. There can be **zero to four** correct answers to each question.

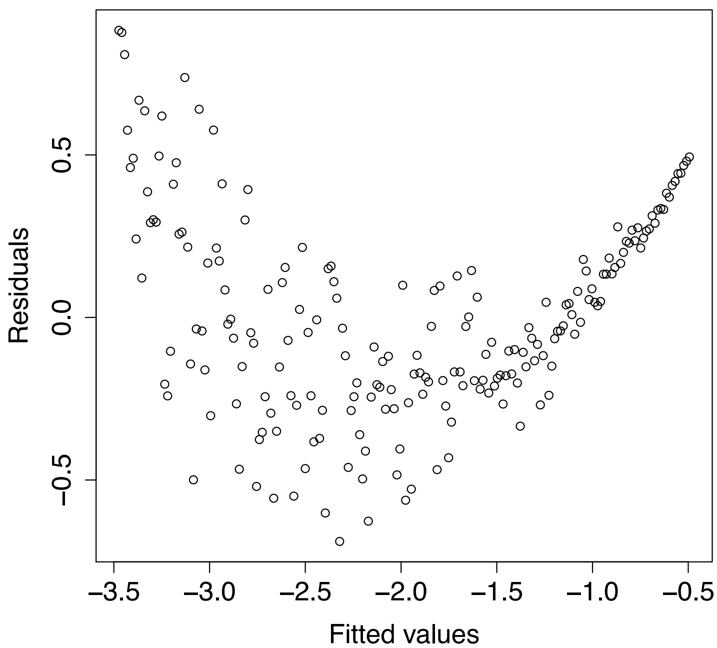
1. For linear regression, the residual plot below indicates that

   ☐ The errors have non-constant variance but the linear assumption is correct.

   ☐ The errors have non-constant variance and the linear assumption is wrong.

   ☐ The linear assumption is correct and the errors have constant variance.

   ☐ The linear assumption is wrong and the errors have constant variance.



2. Interaction/synergy effect of two predictors x_1 and x_2 in regression

   ☐ May be incorporated by including a term of the form βx_1^2.

   ☐ May be incorporated by including a term of the form $\beta x_1 x_2$.

   ☐ May be incorporated by including a term of the form $\beta(x_1 + x_2)$.

   ☐ May be incorporated by using a Generalized Additive Model (GAM) of the form

$$y = \beta_0 + f_1(x_1) + f_2(x_2)$$

.

3. Running a linear regression in **R** and applying the summary function we get the following output

```
Call:
lm(formula = y ~ x + I(x^2))

Residuals:
     Min       1Q   Median       3Q      Max
-2.31384 -0.67054  0.01942  0.62198  2.35304

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01030    0.06014  -0.171    0.864
x            1.02598    0.07512  13.658   <2e-16 ***
I(x^2)       0.07300    0.08409   0.868    0.387
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1

Residual standard error: 0.3079 on 97 degrees of freedom
Multiple R-squared:  0.8892,    Adjusted R-squared:  0.8881
F-statistic: 790.8 on 2 and 97 DF,  p-value: < 2.2e-16
```

Based on this output we can say that

- ☐ The predictor x can be dropped from the linear model since it does not help to predict the response in the presence of the second predictor.
- ☐ The predictor x^2 can be dropped from the linear model since it does not help to predict the response in the presence of the first predictor.
- ☐ Both the predictors x and x^2 can be dropped from the linear model since they have no relationship with the response.

4. The logistic regression model

$$\text{Prob}(Y = 1|X_1, \ldots, X_p) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}},$$

- ☐ always produces a linear decision boundary.
- ☐ can be applied to a classification problem with two classes.
- ☐ cannot be fitted to a dataset in which the observations in different classes are not linearly separable.
- ☐ will become less flexible if some coefficients β_j are forced to be zero.

5. Which of the following are true for the k Nearest Neighbor approach to classification?

- ☐ If we increase k, then it is more likely to overfit.
- ☐ We should choose k to minimize the training error.
- ☐ It does not apply to a 3-class problem if $k < 3$.

6. Which of the following are true for cross-validation?

- ☐ It produces an estimate of the test error.
- ☐ It can be used to choose the value of a tuning parameter for a classification method.

- ☐ For supervised learning problems, it can be used to select a model from several candidate models with different degrees of flexibility.
- ☐ It can be used to choose the degrees of freedom of a cubic spline in a regression problem.

7. Using k-fold cross validation with a dataset of size $n = 100$ to select the value of the budget parameter C in a support vector classifier

- ☐ does not make sense since we should always set $C = 0$ to obtain a separating hyperplane.
- ☐ might give different answers depending on the random splitting.
- ☐ is equivalent to Leave-One-Out Cross Validation when $k = 1$.

8. Comparing the subset $\mathcal{S}_1$ of predictors selected by Lasso and the subset $\mathcal{S}_2$ selected by Best Subset Selection using AIC (on the same dataset),

- ☐ We always have $\mathcal{S}_1$ being a subset of $\mathcal{S}_2$.
- ☐ We always have $\mathcal{S}_1 = \mathcal{S}_2$.
- ☐ A linear model fitted using the predictors in $\mathcal{S}_1$ will always have higher (or equal) training error than $\mathcal{S}_2$.
- ☐ A linear model fitted using the predictors in $\mathcal{S}_1$ will always have higher (or equal) test error than $\mathcal{S}_2$.

9. Consider Lasso and Ridge Regression.

- ☐ We should use Ridge Regression if we want to select a subset of predictors.
- ☐ They always produce the same coefficient estimates if the same tuning parameter λ are used.
- ☐ If $\lambda > 0$, then they correspond to less flexible models compared to the model fitted using the least squares approach.

10. Which of the following operations will typically increase the flexibility of a model?

- ☐ Increasing the parameter λ in a smoothing spline.
- ☐ Increasing the number of knots in a cubic spline.
- ☐ Using a regression tree with more leafs.
- ☐ Pruning a classification tree.

11. Which of the following are true for the Generalized Additive Models (GAMs)?

- ☐ It applies to regression problems.
- ☐ It applies to classification problems.
- ☐ The regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ is a special case of a GAM with predictors (X_1, X_2)
- ☐ The logistic regression model

$$\text{Prob}(Y = 1|X_1) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_1^2}}{1 + e^{\beta_0+\beta_1 X_1+\beta_2 X_1^2}}$$

is a special case of a GAM with the predictor X_1.

12. Which of the following are true for regression trees?

- ☐ The bias typically increases when the number of leafs increases.
- ☐ The bagging approach typically leads to lower variance than a single tree.
- ☐ The number of leafs of a tree will not matter too much as long as it is large enough.
- ☐ A regression tree typically corresponds to a linear model of the response and predictors.

13. Random Forests

- ☐ cannot be applied if some of the predictors are categorical.
- ☐ cannot be applied if the response variable is categorical.
- ☐ with $m = p$ are equivalent to the bagging approach (p is the total number of predictors and m is the number of predictors considered in each split).
- ☐ cannot be applied to a classification problem with more than 2 classes.

14. In the K-NN classification algorithm, using a larger K typically

- ☐ increases the bias.
- ☐ increases the variance.
- ☐ leads to a more flexible model.
- ☐ leads to a lower training error.

15. A Support Vector Machine

- ☐ produces a linear decision boundary if a linear kernel is used.
- ☐ may produce a nonlinear decision boundary is a radial kernel is used.
- ☐ can be applied to a classification problem with more than two classes via the One-Versus-One approach.
- ☐ leads to more flexible models if the value of the budget parameter C is increased.

16. Logistic regression

- ☐ has higher variance than random forests.
- ☐ should always be fitted using all the predictors.
- ☐ produces a classifier with the same number of false positives and false negatives on the training data.
- ☐ has good performance on the training data if the corresponding Area Under the Curve (AUC) is close to one.

17. Using a Generalized Additive Model with smoothing splines when there are p predictors

- ☐ Is always better than using a linear model.
- ☐ Is always better than a regression tree.
- ☐ Is problematic when $p > 3$ since smoothing splines are piecewise cubic curves.

18. Compared to a linear model, a random forest with $B = 500$ trees

☐ Is usually easier to interpret.

☐ Is usually more difficult to interpret.

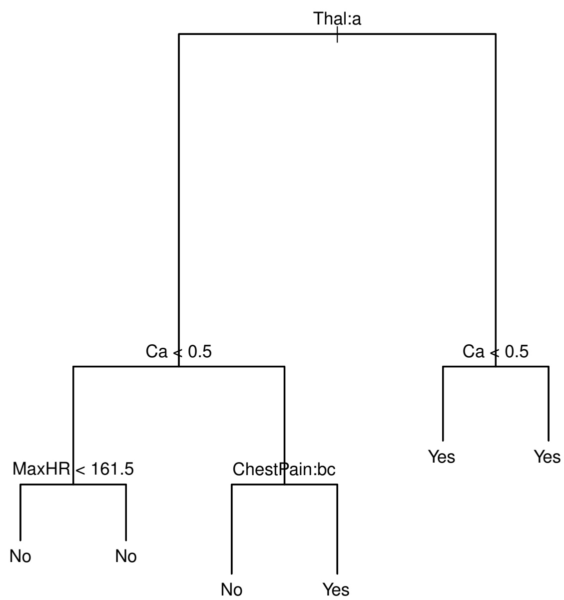
☐ Is usually more flexible.

19. Consider a classification tree depicted in the plot below. Based on that tree, the predicted value $\hat{y}$ for a new observation with predictors Ca=1.2, ChestPain=b, MaxHR=150 and Thal=a will be given by

☐ No.

☐ Yes.

☐ Yes or No depending on the values of the other predictors.

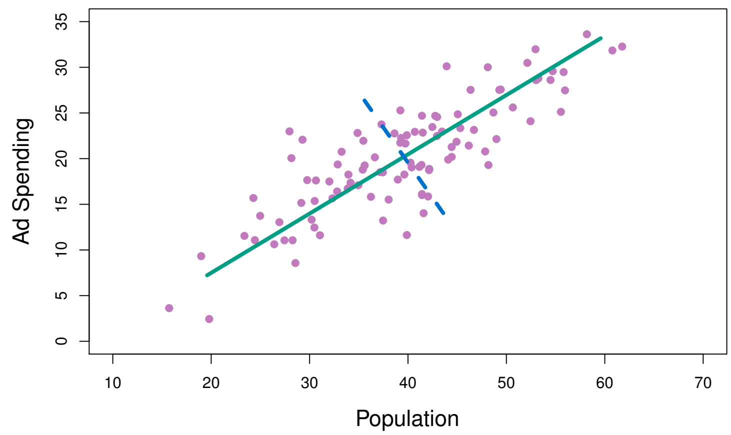


20. Which is/are true for the plot below?

☐ The solid line corresponds to the first principal component of the data.

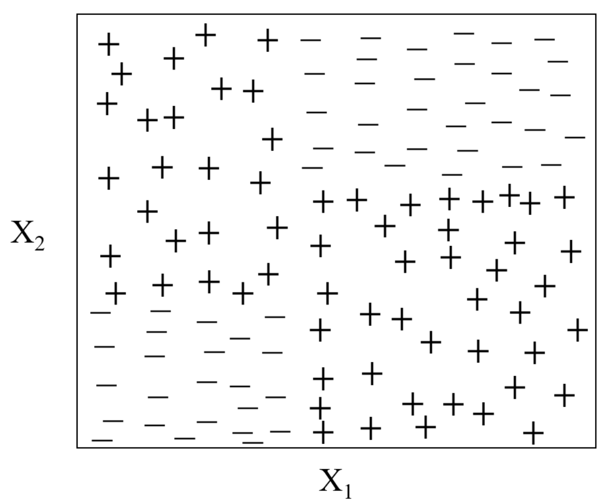
☐ The dashed line corresponds to the first principal component of the data.

☐ Among all straight lines, the solid line has the lowest total squared distance to the observations.



21. Consider a two-dimensional training dataset shown in the plot below, which contains observations in two classes ("+" and "−"). Which of the following methods would you expect to perform best in terms of classification accuracy on a test set generated in the same way?

- ☐ Lasso.
- ☐ Logistic Regression Generalized Additive Models with step functions.
- ☐ Support Vector Machines with linear kernels.
- ☐ Classification trees.



22. Which of the following is true for clustering?

- ☐ The value of k in k-means clustering needs to be chosen by cross-validation.
- ☐ We should choose k in k-means clustering to minimize the total within-cluster variation.
- ☐ The clustering results are usualy not affected by the choice of the distance metric.
- ☐ One may obtain a specific number of clusters by cutting the dendrogram produced by hierarchical clustering.

23. Which of the following are appropriate for selecting a subset of predictors?

- ☐ Principal Component Analysis (PCA).
- ☐ The p-values of the coefficient estimates of a linear regression model.
- ☐ Ridge regression.
- ☐ Lasso.

24. Which of the following can be used to fit a nonliear classifier?

- ☐ Support Vector Machines with appropriate kernels.
- ☐ Logistic Regression Generalized Additive Model.
- ☐ Random Forests.
- ☐ The K Nearest Neighbor algorithm.