Quiz 1, 10 minutes

Question 1: Compute the 2-nn estimator at $x_0 = 2$ based on the data (x_i, y_i) taking values $(0, 2), (2, 5), (3, 5), (5, 2), (6, 2)$.

Question 2: The table below shows 10-fold cv errors and $10^{-1/2}\widehat{se}$ for different k. Which value of k would be selected based on 10-fold cv (**not** using the 1-se rule)?

k	1	2	3	4	5
cv	5	2	3	2.5	5
$10^{-1/2}\widehat{se}$	1	1.5	1.5	2	3

Question 3: Assume you have a sample of size $n = 20$ generated from the model $y_i = f(x_i) + \varepsilon_i$ where ε_i are i.i.d. with variance σ^2 and all x_i are different. Which of the following will always get smaller when k is increased from $k = 5$ to $k = 10$?

1. The squared bias of k-nn regression at a fixed x_0.
2. The variance of k-nn regression at a fixed x_0.
3. The MSE of k-nn regression at a fixed x_0.

Question 4: Which of the following statements are *always* true for k-nn regression with a sample of size $n = 50$?

1. 7-nn regression gives a constant $\widehat{f}$ (i.e. $\widehat{f}(x_0)$ is independent of x_0).
2. 50-nn regression gives a test error of 0.
3. 50-nn regression leads to a training error of 0.

Quiz 2, 10 minutes

1. Consider a linear model with 3 predictors, A, B, C. The following table gives the RSS for each combination of predictors in a linear model.

none	A	B	C	A,B	A,C	B,C	A,B,C
6	5	4	3.5	2	2.5	3	0.5

   1.1 Which candidate model with 1 predictor is selected by backward stepwise selection?
   1.2 Assume that additionally $\hat{\sigma}^2 = 0.1$ and $\log n = 5$. Which model will be selected by best subset selection with AIC?

2. Assume you have 4 predictors. Which statements are true?
   (a) Forward stepwise selection and backward stepwise selection will always select the same model with 1 predictor.
   (b) Forward stepwise selection and best subset selection will always select the same model with 1 predictor.
   (c) Backward stepwise selection and best subset selection will always select the same model with 3 predictors.

3. Comparing model selection based on AIC and BIC when $\log n = 33.2$, which of the following statements are true?
   (1) There are cases when BIC will select models with fewer predictors.
   (2) Sometimes AIC and BIC will select models with the same number predictors.
   (3) AIC and BIC will never select the same model.

Quiz 3, 10 minutes

1. The table below gives values for predictors x and response y

x	0	1	2	3	4	5
y	5	2	1	0	6	-2

   Assume you run a non-linear regression using step functions with intervals $(-2, 3.5], (3.5, 7]$. Compute the corresponding value for $\hat{f}(6)$.

2. Consider a setting where you have $p = 500$ predictors and $n = 100$ observations. Assume none of the predictors are constant. Which of the following methods will give a unique estimator?
   (i) Ridge regression with $\lambda = 10$.
   (ii) Ridge regression with $\lambda = 1$.
   (iii) A linear model with one predictor that was selected by best subset selection.

3. As discussed in lectures piecewise polynomial regression of degree 1 with intervals $(1, 2], (2, 5]$ can be written in the form

$$g(x) = b_1 f_1(x) + ... + b_d f_d(x).$$

   What are the functions $f_1, ..., f_d$ in this example?

4. How many degrees of freedom does a polynomial spline of degree 2 with two knots have (counting the way we counted in lectures)?

Quiz 4, 10 minutes

1. The table below gives values for predictors x and response y

x	0	1	2	3
y	0	2	1	3

   You run a local constant regression with $K(x) = I\{x \in [0, 1]\}$ and span $s = 1$. What is $\hat{f}(1)$?

2. Which of the following will make a model **less flexible**?
   (i) Increasing the degree in local polynomial regression while holding the span fixed.
   (ii) Removing a predictor from a gam model.

3. Which of the following will make a model **more flexible**?
   (a) Choosing a smaller λ in smoothing splines.
   (b) Adding an interaction to a gam model.

4. Which statements are true for smoothing splines?
   (1) If $\lambda = 0$ and all x_i are different we get a training error of zero.
   (2) For very large λ smoothing splines are similar to linear regression.

Quiz 5, 10 minutes

1. Which corresponds to making a model **more flexible**?
   (1) Adding more leafs to a tree.
   (2) Increasing the number of trees in boosting.
   (3) Increasing the interaction depth in boosting while holding everything else fixed.

2. Which statements are true for boosting?
   (a) Choosing an interaction depth of 1 will always lead to the best predictions.
   (b) A larger learning rate usually leads to better predictions.

3. Which of the following can lead to overfitting?
   (i) Too many trees in a random forest.
   (ii) Too many trees in boosting.

4. Consider the regression tree on the blackboard. What is your prediction for a new observation with $X1 = 5, X2 = 3$?

Solutions

Quiz 1
Q1: (5+5)/2 Q2 k=2 Q3 (2) Q4 none

Quiz 2:
Q1.1 B Q1.2 A,B,C Q2 (b), (c) Q3 (1), (2)

Quiz 3:
Q1: (6-2)/2 Q2 (i), (ii), (iii)
Q3 $f_1(x) = I\{x \in (1,2]\}, f_2(x) = xI\{x \in (1,2]\}, f_3(x) = I\{x \in (2,5]\}$
$f_4(x) = xI\{x \in (2,5]\}$
Q4 3+2

Quiz 4:
Q1: that is just the 4-nn estimator, so the solution is $(0+2+1+3)/4$
Q2: (ii), Q3: (a), (b), Q4: (1), (2)

Quiz 5:
Q1: (1), (2), (3) Q2: none Q3: (ii)