NAME (PRINT): ______________________

Last/Surname First /Given Name

STUDENT #: ______________ **SIGNATURE:** ______________

UNIVERSITY OF TORONTO MISSISSAUGA
DECEMBER 2017 FINAL EXAMINATION
STA312H5F
Topics in Statistics: Applied Statistical Modelling
Stanislav Volgushev
Duration - 2 hours
Aids: None

The University of Toronto Mississauga and you, as a student, share a commitment to academic integrity. You are reminded that you may be charged with an academic offence for possessing any unauthorized aids during the writing of an exam. Clear, sealable, plastic bags have been provided for all electronic devices with storage, including but not limited to: cell phones, SMART devices, tablets, laptops, calculators, and MP3 players. Please turn off all devices, seal them in the bag provided, and place the bag under your desk for the duration of the examination. You will not be able to touch the bag or its contents until the exam is over.

If, during an exam, any of these items are found on your person or in the area of your desk other than in the clear, sealable, plastic bag, you may be charged with an academic offence. A typical penalty for an academic offence may cause you to fail the course.

Please note, once this exam has begun, you ***CANNOT*** *re-write it.*

Marks possible: 45 **Marks achieved**:

Problem	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Possible	1.5	1.5	3	1	1	1	2	3	1	4	5	3	2	3	2.5	3	1	1	1	1	3.5
Achieved																					

1. (**0.5 marks each**) Assume that `D` is a data frame that contains 4 columns with names `Y`, `X`, `V`, `Z`. For each of the following specifications, write down the regression function that corresponds to the `lm` call in **R**. Example:

```
lm( Y ~ V, data = D)
```

corresponds to $f(x, v, z) = b_1 + b_2 v$.

   (a) `lm( Y ~ . - X, data = D)`

   (b) `lm( Y ~ V +  V:Z, data = D)`

   (c) `lm( Y ~ I(X^2) + I(V^2) + Z, data = D)`

2. (**0.5 marks each**) Assume that `D` is a data frame that contains 4 columns with names `Y`, `X`, `V`, `Z`. For each of the following regression functions, decide if they can be formulated as a linear regression in **R** (here, $b_1, ..., b_3$ are unknown). If yes, write the **R** call you would use.

   (a) $f(x, v, z) = b_1 + b_2 \sin(x + b_3)$

   (b) $f(x, v, z) = b_1 + b_2 z + b_3 x^2$

   (c) $f(x, v, z) = b_1 + b_2 xv + b_3 x^7$

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

```
Call:
lm(formula = y ~ x1 + x2 + I(x2^2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.29881 -0.07232 -0.01262  0.07602  0.28263

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.010000   0.011326  -1.048    0.297
x1          -0.010000   0.011095  -0.762    0.448
x2           0.900000   0.011910  83.384   <2e-16 ***
I(x2^2)      1.200000   0.011635  86.359   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.111 on 96 degrees of freedom
Multiple R-squared:  0.9929,    Adjusted R-squared:  0.9927
F-statistic:  4487 on 3 and 96 DF,  p-value: < 2.2e-16
```

(a) (**1 mark**) Given the output above, what is your prediction for a new observation with predictor values `x1 = 1, x2 = 2`? You don't need to simplify your answer, it is enough if you write down the correct formula.

(b) (**0.5 marks**) What is the value for R^2 in the model above?

(c) (**1.5 marks**) Given the output above, would it make sense to consider a linear model without the predictor $x1$? Why/why not?

4. (**1 mark**) Assume that you observe data (x_i, y_i) with values $(1,2),(2,1),(3,3),(4,7)$. Compute the 2-nn estimator for $x = 5/4$.

5. (**1 mark**) Which output will running the following code in **R** give?

```
x = array(0,3)
for(i in 1:length(x)){
 x[i] = i -2
}
x[c(1,3)]
```

6. (**1 mark**) Which output will running the following code in **R** give?

```
x = 0
for(i in 1:3){
 x = i-1
}
x
```

7. (**2 marks**) Consider the data set with observations (x_i, y_i) given by $(1,2),(3,4),(1.5,5)$, $(3.5,2),(4,0)$ (here y_i are outcomes and $x_i \in R$ are predictors). If you fit a model using a step function with intervals $(0,2],(2,5]$ as regression function, what is your prediction for a new observation with predictor $x = 3$? Show your computations to get full marks.

8. The table below provides values of estimated coefficients in a linear model of the form

$$f(x) = b_1 + b_2 x$$

for either lasso or ridge regression for different values of λ (you will need to figure out which one). All estimators were computed on the same data set.

λ	0	1	2	3	10	20
$\widehat{b_1}$	1	1	1	1	1	1
$\widehat{b_2}$	2	1.5	1	0	0	0

(a) (**1 mark**) Based on this table, what is the usual least squares estimator for b_1, b_2 for this data set?

(b) (**2 marks**) Is this lasso or ridge regression (it is one of the two)? Justify your answer.

9. (**1 mark**) The table below gives estimated MSE values (vie 5-fold cross-validation) for k-nn regression with different values of k. Which of the k in the table would you select based on cross-validation?

k	1	2	3	4	5	6	7
CV error	10	8	1.5	3	1	2.5	5

10. Assume you want to model the influence of a predictor `x` on a response `y` by the regression function f which is a cubic spline (i.e. polynomial spline of degree 3) with knots in the points 0 and 1.

   (a) (**1 mark**) How many degrees of freedom (**counting the way we counted in lectures**) does this model have?

   (b) (**1 mark**) Assume that `D` is data frame with columns `y` (response) and `x` (predictor). Which `R` input would you use to fit this model?

   (c) (**2 marks**) Write down functions $g_1, g_2, ..., g_d$ such that f is a a cubic spline with knots in the points 0 and 1 if and only if

$$f(x) = b_1 + b_2 g_1(x) + ... + b_{d+1} g_d(x)$$

   for some $b_1, ..., b_{d+1} \in R$.

11. 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.

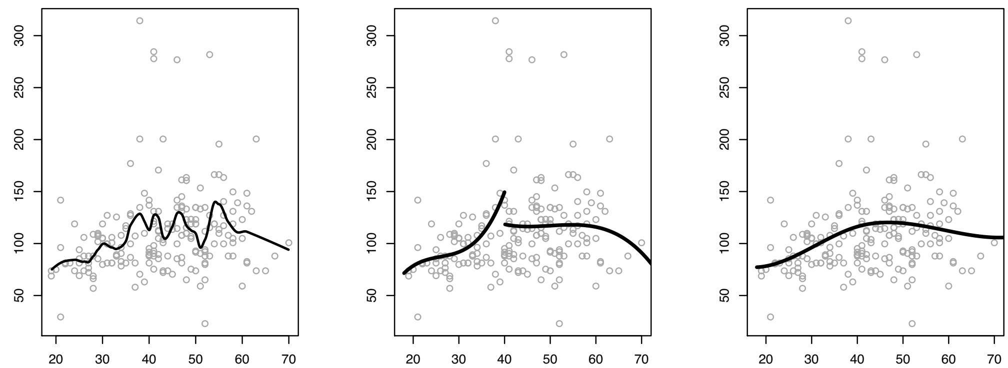
none	A	B	C	A,B	A,C	B,C	A,B,C
6	4	4.5	3	1	2	2.5	0.5

(a) (**1 mark**) Which model with 2 predictors is selected by best subset selection?

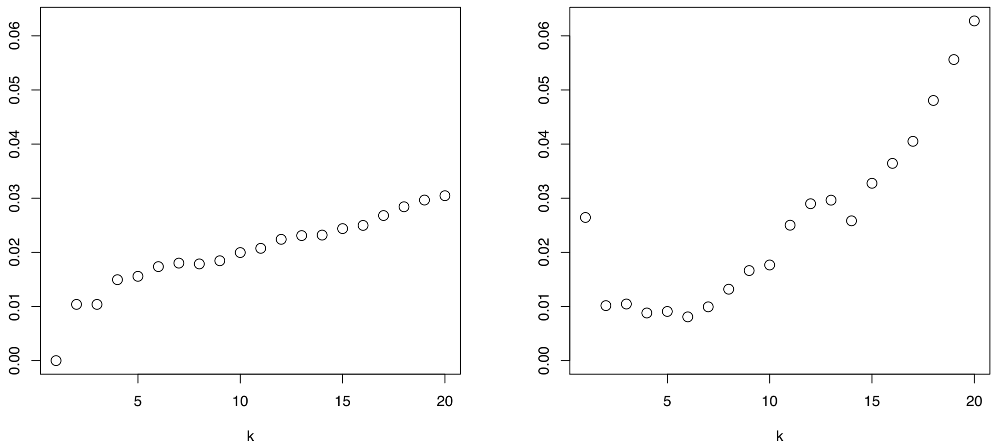
(b) (**1 mark**) Which model with 2 predictors is selected by forward stepwise selection?

(c) (**1 mark**) Which model with one predictor is selected by backward stepwise selection?

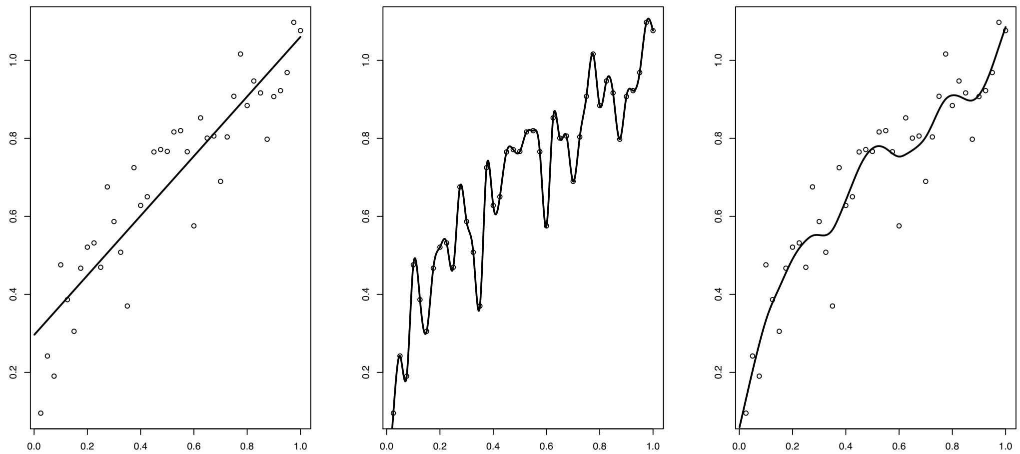
(d) (**2 marks**) Assume that additionally $\hat{\sigma}^2 = 1$ and $\log n = 10$. Which model will be selected by best subset selection with BIC penalty? Show your computations to get full marks.



12. (**3 marks**) The plots at the top of this page show the graphs of 3 different regression functions: a spline of degree 3 with 1 knot, local regression, and a piecewise polynomial of degree 3 with one knot. All functions were fitted to the displayed data set by least squares. Which plot is which? Justify your answer.

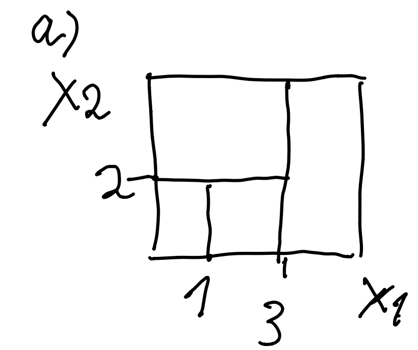


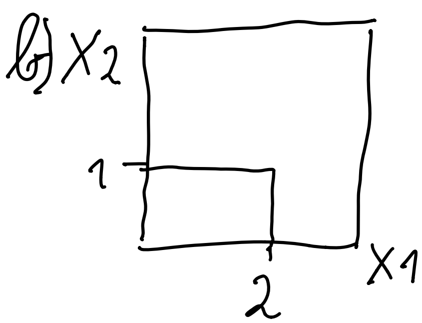
13. (**2 marks**) The plots above show the test and training MSE for k-nn regression as a function of k. Which one is which? Justify your answer.



14. (**3 marks**) The plots above show the training set (dots, same data in all plots) and smoothing splines corresponding to $\lambda = 0$, $\lambda = 1$, $\lambda = 100$. Which plot corresponds to which smoothing spline? Justify your answer.

15. (**1.5 marks each**) For each of the two partitions of the predictor space given below, is it possible to represent them as a regression tree? If yes, draw the corresponding tree.

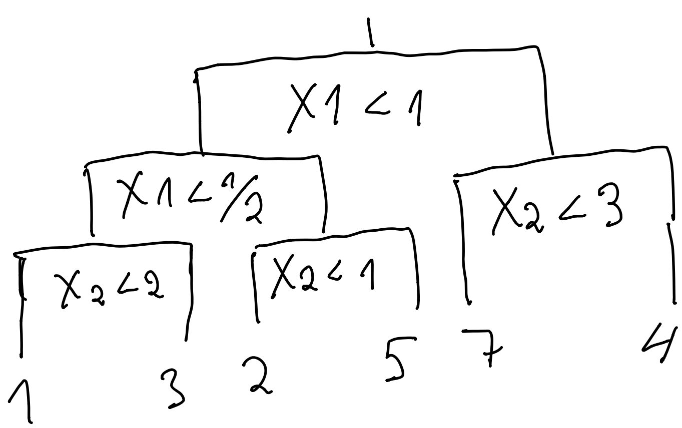


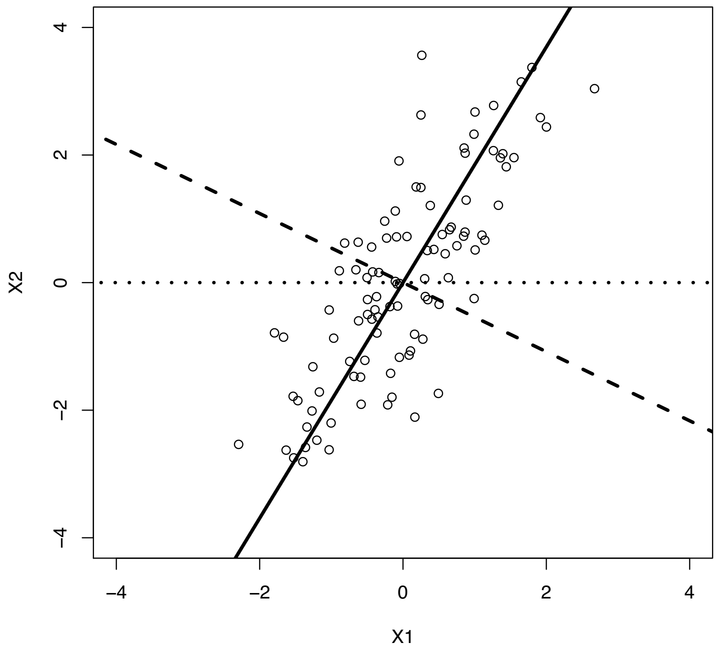


16. For the regression tree shown below

    (a) (**1 mark**) What is the predicted value for a new point with predictor $X = (1/4, 1)$?

    (b) (**1.5 marks**) Draw the corresponding partition of the predictor space.





17. (**1 mark**) Which of the lines in the plot above corresponds to the first principal component in PCR? (The points correspond to predictors, predictors are two-dimensional with components $X1, X2$)

18. (**1 mark if and only if ticks are only next to correct answers. Any number of answers (0-3) can be correct**) Which of the following methods can still give reasonable predictions when the number of predictors p is larger than the number of observations n?

   - ☐ PCR with number of principal components smaller than n
   - ☐ The usual least squares estimator with all predictors.
   - ☐ Ridge regression with suitably chosen value for λ.

19. (**1 mark if and only if ticks are only next to correct answers. Any number of answers (0-3) can be correct**) Compared to growing a large single tree, using bagging

   - ☐ Can be better because it helps to reduce variance.
   - ☐ Will never be better.
   - ☐ Will only be better if the number of trees is small.

20. (**1 mark if and only if ticks are only next to correct answers. Any number of answers (0-3) can be correct**) By pruning large trees we try to

   - ☐ Reduce the bias.
   - ☐ Reduce the variance.
   - ☐ Obtain a more flexible model.

21. (**Check the boxes next to correct statements. 0.5 marks for each correct box**) Which of the following changes correspond to *increasing* the flexibility of a model (usually this corresponds to decreasing the test error)

   - ☐ Removing predictors from a linear model.
   - ☐ Increasing λ in lasso.
   - ☐ Decreasing the span in local regression.
   - ☐ Pruning back large regression trees.
   - ☐ Adding knots to a polynomial spline.
   - ☐ Increasing the degree of a polynomial spline while holding the knots fixed.
   - ☐ Choosing smaller k in k-nn regression.

END OF EXAM