Midterm for STA 314: partial solutions

October 25, 2019

There can be zero to four correct answers to each multiple choice question. One point is assigned to a multiple choice question if and only if all boxes next to the correct answers to this question are checked and no box next to an incorrect answer to this question is checked. For all other questions, if not stated differently, one point is assigned if the task is completely and correctly fulfilled. No negative points are assigned for wrong answers. Throughout, we use the notation, models, estimators, etc. introduced in the lecture. Good luck!

Time: **90 minutes.**

Aids allowed: none.

Name:

Studet ID:

1. (0.5 marks each) *Assume that* `D` *is a data frame that contains 4 columns with names* `Y`, `X`, `V`, `W`. *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 ~ . - V + W:X, data = D)`

   (b) `lm( Y ~ I(sin(X)) + I(V^3), 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) $f(x, v, z) = b_1 + b_2 zv + b_3 \cos(x^2)$

   (b) $f(x, v, z) = b_1 + b_2 e^{z+b_3 v}$

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

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

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) -1.100000   0.011326  -1.048    0.297
x1           0.900000   0.011910  83.384   <2e-16 ***
x2          -0.050000   0.011095  -0.762    0.448
x1:x2        0.100000   0.011635  0.359     0.672
---
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:  0.7929,    Adjusted R-squared:  0.7927
F-statistic:  4487 on 3 and 96 DF,  p-value: < 2.2e-16
```

(a) (0.5 marks) What is your prediction for a new value with predictors $x1 = 2, x2 = 3$? (you don't need to simplify your answer, just write down the correct formula)

$-1.1 + 0.9 * 1 - 0.05 * 3 + 0.1 * 1 * 3$

(b) (0.5 marks) What is the value for R^2 in the model above?
0.7927

4. (1 mark) *Assume that you observe data* (x_i, y_i) *with values* $(1,2),(2,3),(4,5),(0,2),(3,7)$. *Compute the 3-nn estimator for* $x = 10$.

   $(3+5+7)/3$

5. *You run 12-fold cross-validation on a data set and obtain the following values for the cross-validated error and standard error for different numbers of predictors in a linear model.*

number predictors	0	1	2	3	4	5	6	7	8	9
cv error	12	8	6	2	3	8	2	6	1	30
$12^{-1/2}\widehat{se}$	5	4	3	3	5	6	7	5	6	5

   (a) (0.5 marks) Which number of predictors would you select based on 12-fold cross-validation?

   8

   (b) (1 mark) Which number of predictors would you select based on the one standard error rule?

   2

6. *Which output will running the following code in **R** give?*

```
x = array(0,2)
for(k in 1:length(x)){
 x[k] = 5 + k
}
min(x)
```

7. *Which output will running the following code in **R** give?*

```
set.seed(1)
x = rnorm(1)
set.seed(1)
y = rnorm(1)
x-y
```

8. *Qualitative predictors in linear regression*

   X Can be incorporated using dummy variables.

   ☐ Can not be incorporated if the qualitative predictor takes two different values.

   ☐ Can only be incorporated if the qualitative predictor takes two different values.

9. *Adding an interaction effect to a model with two predictors will.*

   ☐ Usually increase the training error.

   X Usually decrease the training error.

   ☐ Never change the training error.

10. *A small value of R^2 (a value close to zero) in a linear regression model*

    ☐ Means that the regression model is incorrect.

    X Means that the model does not help to obtain a prediction that is much better than the sample mean.

    ☐ Means that the relationship between the predictors and the response is non-linear.

    ☐ Means that there is no relationship between the predictors and the response.

11. *Removing predictors from a linear regression model*

    - ☐ Will never increase the test error.
    - X Will never decrease the training error.
    - ☐ Will never affect R^2.

12. *Assume you have 5 candidate predictors. Denote by* M_{best}, M_F, M_{back} *candidate models with two predictors chosen by best subset, forward stepwise and backward stepwise selection respectively. Which of the following are true?*

    - X Always: $RSS(M_{best}) \leq RSS(M_F)$
    - ☐ Always: $RSS(M_{best}) = RSS(M_F)$
    - X Always: $RSS(M_{best}) \leq RSS(M_{back})$

13. *Comparing model selection (after using best subset to select candidate models) based on AIC and BIC for* $\log n > 10$

    - ☐ AIC and BIC will never select the same model.
    - ☐ BIC will always select models with strictly fewer predictors.
    - ☐ AIC will always select models with strictly more predictors.

14. *Which statements are true for ridge regression?*

    - ☐ Large values of λ can lead to coefficients being set to exactly zero.
    - ☐ Increasing λ will decrease the bias of predictions.
    - ☐ Increasing λ will increase the variance of predictions.

15. *Using* 10*-fold cross validation with a data set of size* $n = 10$ *to select a tuning parameter*

    - X Will always result the same tuning parameter since it does not involve any randomness.
    - ☐ Might give different answers depending on the randomness in the split.
    - ☐ Will not work since $n = 10$.

16. *Which of the following methods can be used for model selection (i.e. for selecting relevant predictors)?*

    X LASSO

    ☐ Ridge regression

    X Best subset selection with BIC.

17. (1.5 marks) Fill out the table below with numbers such that best subset and forward stepwise selection choose different candidate models with two predictors. Be careful to only provide numbers which would correspond to possible values for the training error.

Predictors	None	A	B	C	A,B	A,C	B,C	A,B,C
Training error	7	6	5	4	2	2.5	3	1

18. (2 marks) Consider a model of the form $y_i = f(x_i) + \varepsilon_i$, where ε_i are i.i.d. with $E[\varepsilon_i] = 0, \mathrm{Var}(\varepsilon_i) = 2$ and assume that $\widehat{f}(x_0)$ is an estimator that you obtained on the training set. Consider a new observation $y_0 = f(x_0) + \varepsilon_0$ where x_0 is fixed and ε_0 is independent of $\widehat{f}(x_0)$ and satisfies $E[\varepsilon_0] = 0$. Is it possible that $E[(\widehat{f}(x_0) - f(x_0))^2] = 1$? Justify your answer to get full marks.

Yes, this is possible. Example: k-nn with $f(x) = 0$ for all x, $k = 2$.

19. (2 marks) Write down the definition of RSS and TSS in a linear model with a one-dimensional predictor x. Prove that $TSS \geq RSS(\hat{b})$ where $\hat{b}$ is the least squares estimator (the estimator we discussed in class) in the linear model without using properties of R^2 discussed in lectures.

See lectures

20. (3 marks) Consider a data set (x_i, y_i) with fixed one-dimensional predictors x_i with $\bar{x} := n^{-1} \sum_{i=1}^n x_i = 0$ and $y_i = \varepsilon_i$ where ε_i are i.i.d. normal random variables with mean zero and variance 1. Consider two candidate models: one with predictors x_i and one with no predictors. Prove: the probability that BIC selects a model with no predictors tends to one as $n \to \infty$. You may assume that $\sigma^2 = 1$ in BIC is a fixed number.

Hint You may use the following formulas which we proved in lectures for the model above: in a linear regression model with predictors x_i the coefficients take the form $\hat{b}_1 = n^{-1} \sum_{i=1}^n y_i$, $\hat{b}_2 = (n^{-1} \sum_{i=1}^n x_i y_i)/(n^{-1} \sum_{i=1}^n x_i^2)$.

See lectures

Empty page for additional calculations.