Quiz 3

You have 30 minutes to complete this 10-question quiz. The questions, a mix of multiple choice, fill-in-the-blank, and numeric answers, are weighted equally. You can consult any course materials, the internet, or R. However, you must complete the quiz individually.

All 10 questions relate to the crime data discussed in Unit 3. These data contain 97 socioeconomic factors and a response violentcrimes.perpop for 90 communities in Florida, 72 of which are used for training. The training mean and standard deviation of the feature pct.kids.nvrmarried (percentage of kids whose parents were never married) and of the response violentcrimes.perpop (violent crimes per capita) are given in the table below.

We have two communities (A and B) for which we'd like to make predictions, which have exactly the same set of features except that Community A has pct.kids.nvrmarried = 4 while Community B has pct.kids.nvrmarried = 10.4.

Variable	Training mean	Training std. deviation	Community A	Community B
pct.kids.nvrmarried	4	3.2	4	10.4
violentcrimes.perpop	1159	826.9	?	?

Stimulus

Lasso regression
We run a 10-fold cross-validated lasso regression on the training set; below are the trace plot and CV plot. The vertical dashed line in the trace plot corresponds to λ chosen via the one-standard error rule.

Lasso trace plot

---

1 Fill in the Blank   0.5 points

The predictions for lambda = 1000 perform poorly due to high

> bias

;

the predictions for lambda = 10 perform poorly due to high

> variance
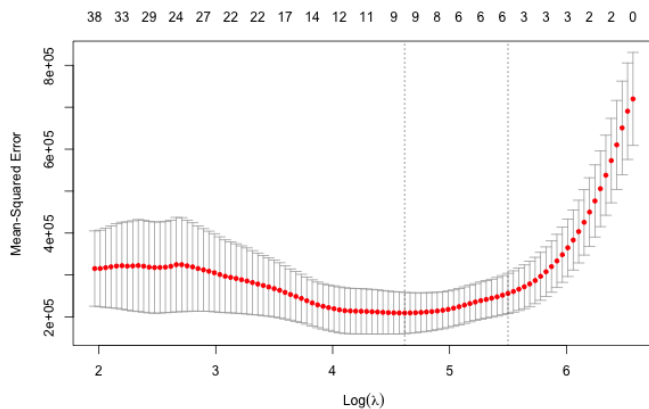
.

> variance   > bias

num.kids.nvrmarried — pct.kids.nvrmarried — pct.people.dense.hh
pct.pop.underpov — pct.youngkids2parents
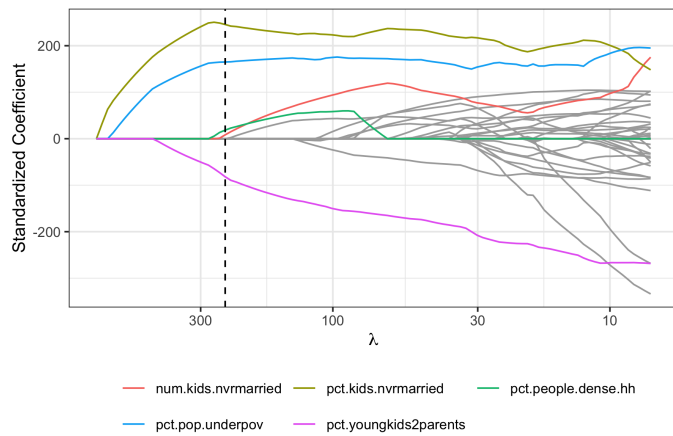
Lasso CV plot

2 Numeric 0.5 points

What is the number of features with nonzero coefficients in the model minimizing the CV error? (Please exclude the intercept.)

9

3 Fill in the Blank 0.5 points

Based on the lasso with lambda = 10, the predicted violentcrimes.perpop for Community A is

less than

the predicted violentcrimes.perpop for Community B by

400

.

(Express your answer numerically, rounded to the nearest 10).

⋮⋮ greater than

4 Multiple Choice   0.5 points

Suppose the elastic net were applied with $\alpha = 0.5$ and the same λ as was chosen by the one-standard-error rule for lasso regression. Which of the choices below would be the most likely number of features with nonzero coefficients? (Please exclude the intercept.)

- ○ 3
- ○ 5
- ◉ 15
- ○ 97

---

5 Fill in the Blank   0.5 points

The lasso solution with lambda = 0

| is undefined |

.

| has 38 features with nonzero coefficients |

| has 97 features with nonzero coefficients |

| has 0 features with nonzero coefficients |

Logistic lasso regression

We transform violentcrimes.perpop to a binary response (1="higher-crime" if violentcrimes.perpop > 1089 and 0="lower-crime" otherwise), where 1089 is the median of violentcrimes.perpop in the training data. We run a 10-fold cross-validated logistic lasso regression on the training set with this binarized response; below are the trace plot and CV plot.

- The vertical dashed line in the trace plot corresponds to λ chosen via the one-standard error rule.
- You may assume that the two classes of responses are balanced in the test set.

Logistic lasso trace plot

pct.kids.nvrmarried — pct.kids2parents
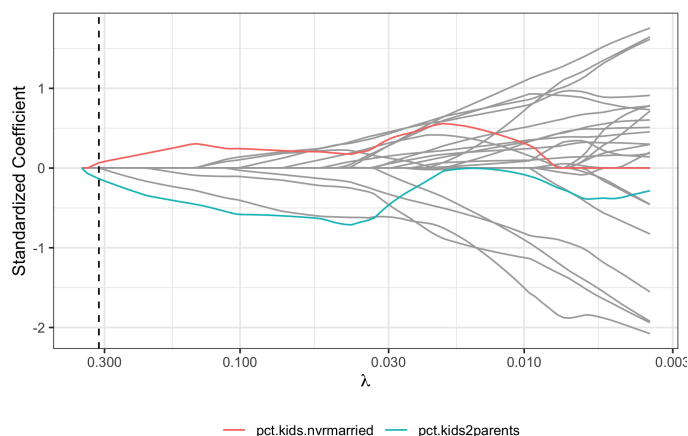
Logistic lasso CV plot

---
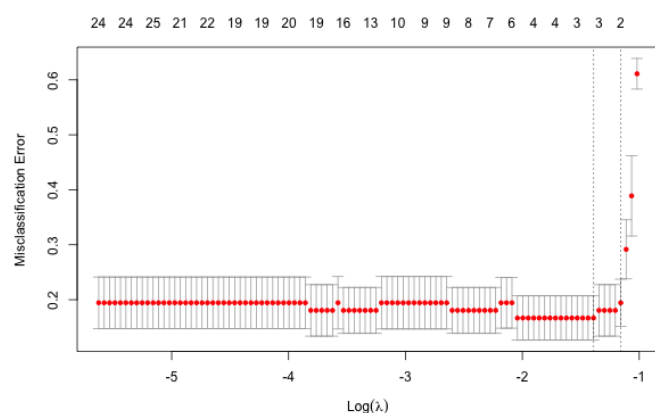
6 Numeric   0.5 points

You build a model for higher-crime probability based on the logistic lasso with $\lambda = 0.1$. The odds that Community A is higher-crime is approximately what factor times the odds that Community B is higher-crime? (Express your answer as a decimal, rounded to the nearest tenths place.)

> 0.6

---

7 Multiple Answer   0.5 points

You build a classifier based on the logistic lasso, with λ chosen based on the one-standard error rule. Which of the following are possible predictions for Communities A and B?

- ☑ Communities A and B are predicted to be higher-crime.
- ☑ Communities A and B are predicted to be lower-crime.
- ☐ Community A is predicted to be higher-crime and Community B is predicted to be lower-crime.
- ☑ Community A is predicted to be lower-crime and Community B is predicted to be higher-crime.

8   Fill in the Blank   0.5 points

You build a classifier based on the logistic lasso, with λ chosen based on the one-standard error rule.

According to the CV estimate, the resulting test misclassification error is about

| 40 | % |

of the test misclassification error of the naive classifier that guesses each class with equal probability. (Round your answer to the nearest 10%).

---

9   Multiple Choice   0.5 points

You build a classifier based on the logistic lasso with $\lambda = 1$. Which of the following numbers is closest to the test misclassification error of the resulting classifier?

- ○   0
- ○   0.25
- ●   0.5
- ○   0.75
- ○   1

Multiple Choice   0.5 points

Each value of λ gives rise to a set of logistic lasso coefficients β, pictured in the trace plot. Each set of coefficients β gives rise to a likelihood. Which value of λ leads to the highest likelihood?

○   0.3

○   0.1

○   0.03

◉   0.01