Midterm 1

Due Oct 15 at 9am **Points** 52 **Questions** 17 **Available** after Oct 13 at 9am
Time Limit 1440 Minutes

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	1382 minutes	7 out of 52 *

* Some questions not yet graded

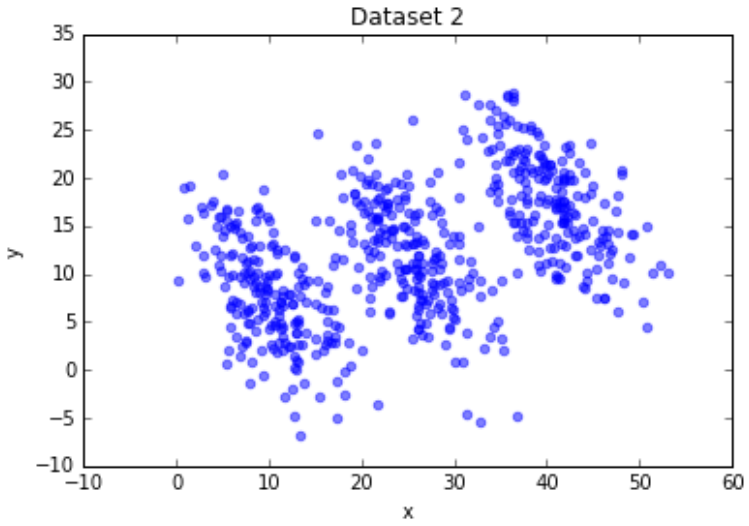
⚠ Correct answers will be available on Oct 17 at 1pm.

Score for this quiz: **7** out of 52 *
Submitted Oct 15 at 8:23am
This attempt took 1382 minutes.

Question 1

1 / 1 pts

From the following dataset, we randomly remove a set of x-values, $X_{test} = x_1, ..., x_n$, and corresponding y-values, $y_{test} = y_1, ..., y_n$.



We use the remaining dataset to train two regression models, a KNN model for $k = 3$ and a simple linear regression model. Which model would you expect to have a lower MSE on the testing set?

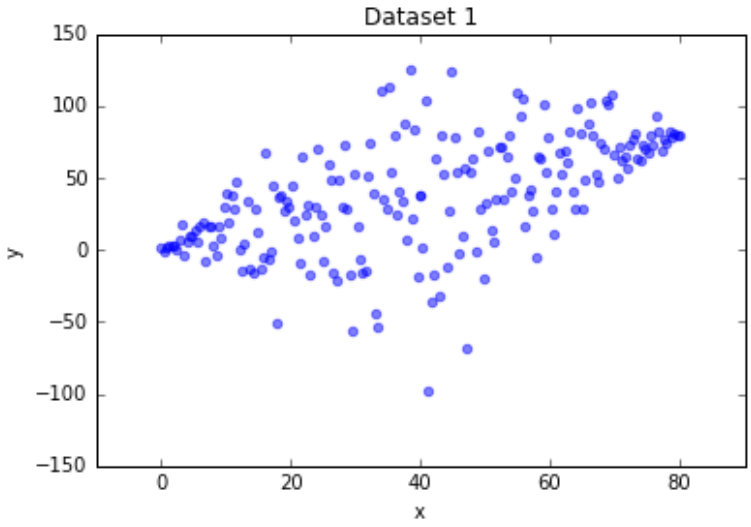
- ○ Simple Linear Regression
- ◉ KNN, with k=3

Their performance will effectively be the same

Question 2

1 / 1 pts

From the following dataset, we randomly remove a set of x-values, $X_{test} = x_1, ..., x_n$, and corresponding y-values, $y_{test} = y_1, ..., y_n$.



We use the remaining dataset to train two regression models, a KNN model for $k = 3$ and a simple linear regression model. Which model would you expect to have a lower MSE on the testing set?

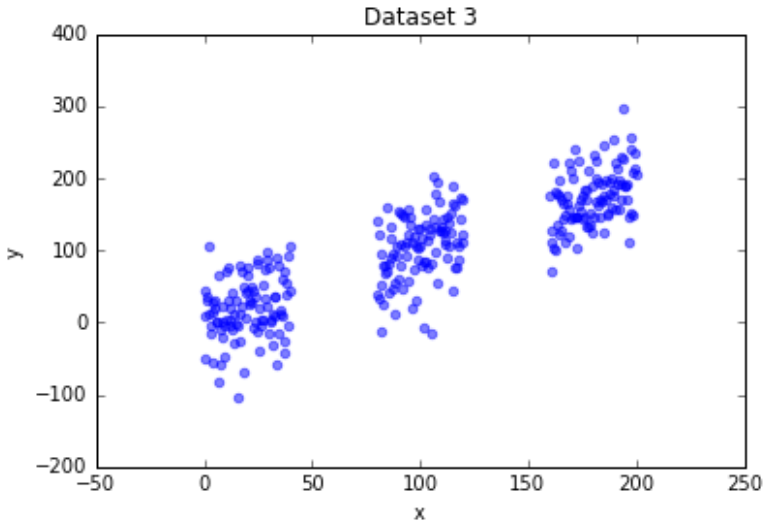
- (•) Simple Linear Regression
- ( ) KNN, with k=3
- ( ) Their performance will effectively be the same

Incorrect

Question 3

0 / 1 pts

From the following dataset, we randomly remove a set of x-values, $X_{test} = x_1, ..., x_n$, and corresponding y-values, $y_{test} = y_1, ..., y_n$.



We use the remaining dataset to train two regression models, a KNN model for $k = 3$ and a simple linear regression model. Which model would you expect to have a lower MSE on the testing set?

- Simple Linear Regression
- KNN, with k=3
- Their performance will effectively be the same

Question 4

1 / 1 pts

Suppose you are given a data set containing medical data for 20 patients. This data set contains the patients' sex, height and weight. There are 10 males and 10 females. You posit that there is a polynomial relationship between height and weight, and that this relationship is different for males and females.

Which of the following visualizations would be most appropriate to intuitively test your hypothesis?

- A bubble chart: each patient is visualized as a bubble over a horizontal height axis, the color of the bubble encodes sex and the radius of the bubble encodes weight.
- Histograms: plot the histogram of male and female heights in plot #1; plot the histogram of male and female weights in plot #2
-

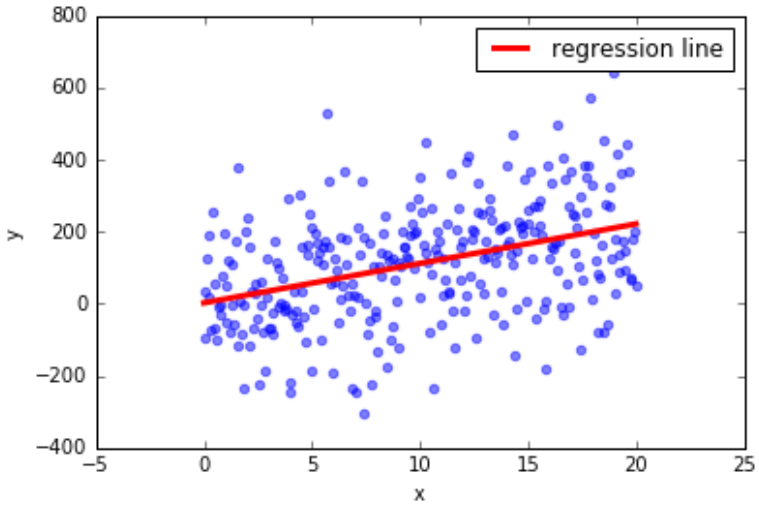
Line plot: order the data by weight; plot the weight and height as points; connect consecutive points with lines; color code lines by sex

Stacked bar plot: pair up males and females with similar weights; plot each pair as a bar with height equal to the female height on top of a bar whose height is equal to the male height; color code the stacked bars by sex

Question 5

1 / 1 pts

The follow simple linear regression model has an R^2 value of 0.1574.



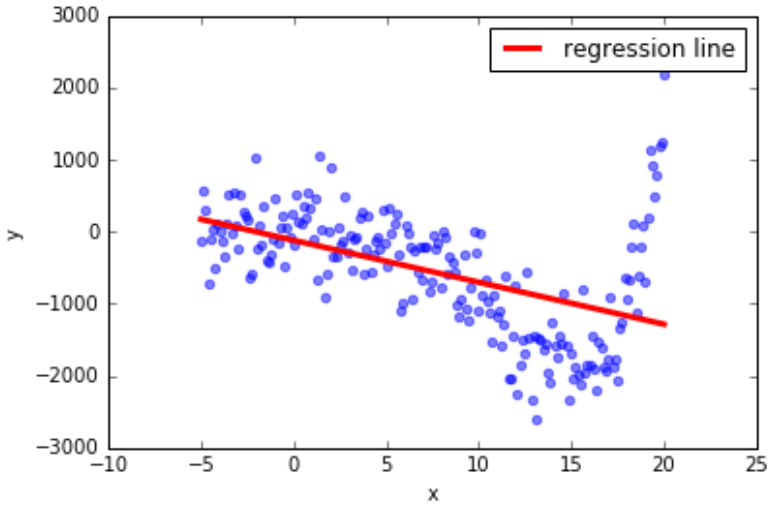
The low R^2 is most likely due to:

- (•) Irreducible error
- ( ) Reducible error
- ( ) Overfitting due to lack of regularization
- ( ) The lack of cross-validation before training the model

Question 6

1 / 1 pts

The follow simple linear regression model has an R^2 value of 0.2469.



The low R^2 is most likely due to:

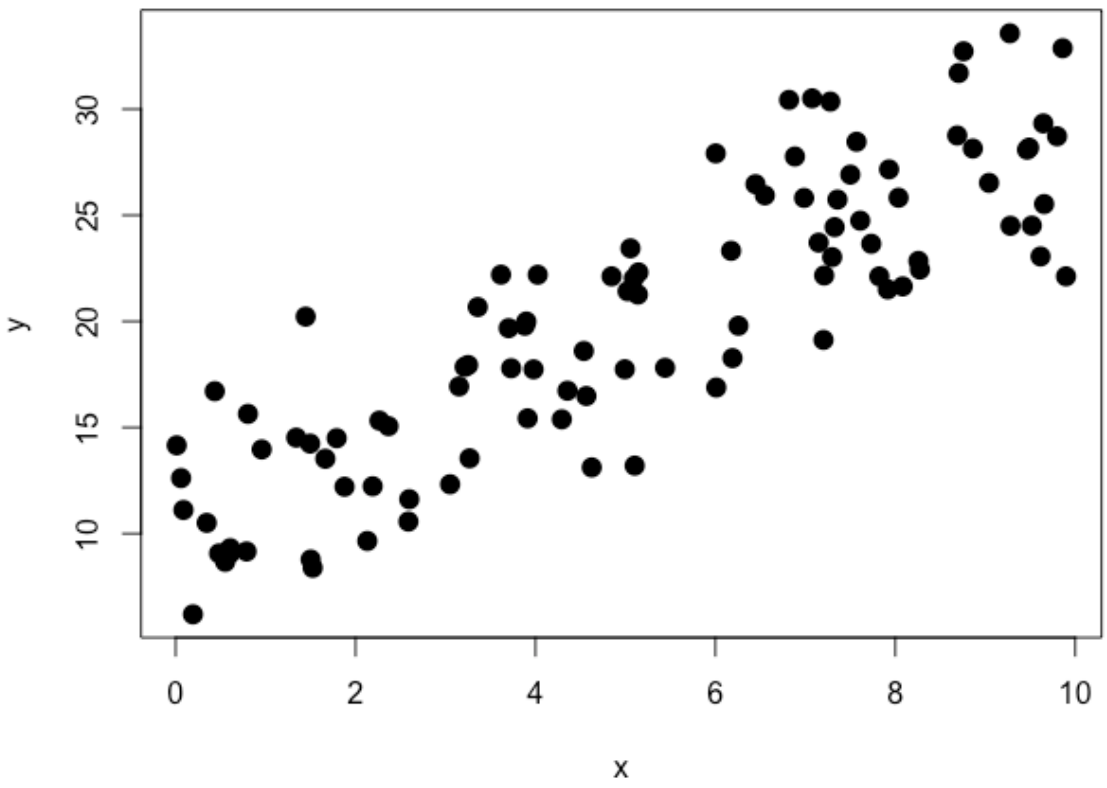
- Irreducible error
- Reducible error
- Overfitting due to lack of regularization
- The lack of cross-validation before training the model

Incorrect

Question 7

0 / 1 pts

Several Regression Models were considered for the following scatterplot of points. Note: the observed mean of y is 20 and the observed mean of x is 5. Provide a reasonable value of R^2 for each of the 4 models below.



Provide a reasonable value for R^2 if the model used was $\hat{y} = 30 - 2x$.

0.2800

Incorrect

Question 8 **0 / 1 pts**

Provide a reasonable value for R^2 if the model used was $\hat{y} = 10 + 2x$.

0.0300

Incorrect

Question 9 **0 / 1 pts**

Provide a reasonable value for R^2 if the model used was $\hat{y} = 20$.

0.7800

Incorrect

Question 10

0 / 1 pts

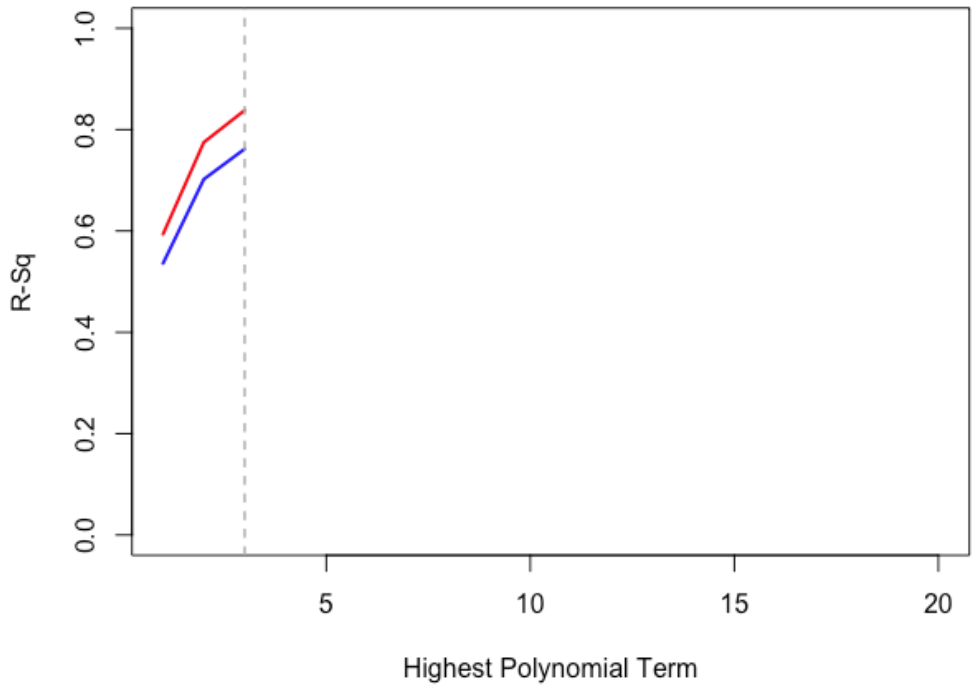
Provide a reasonable value for R^2 if the model used was $\hat{y} = 15 + x$.

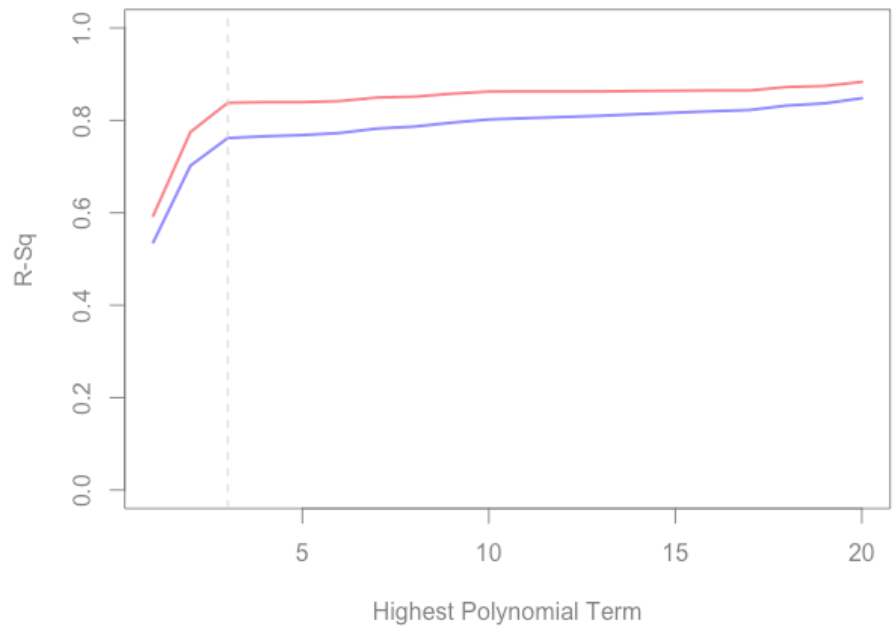
0.0000

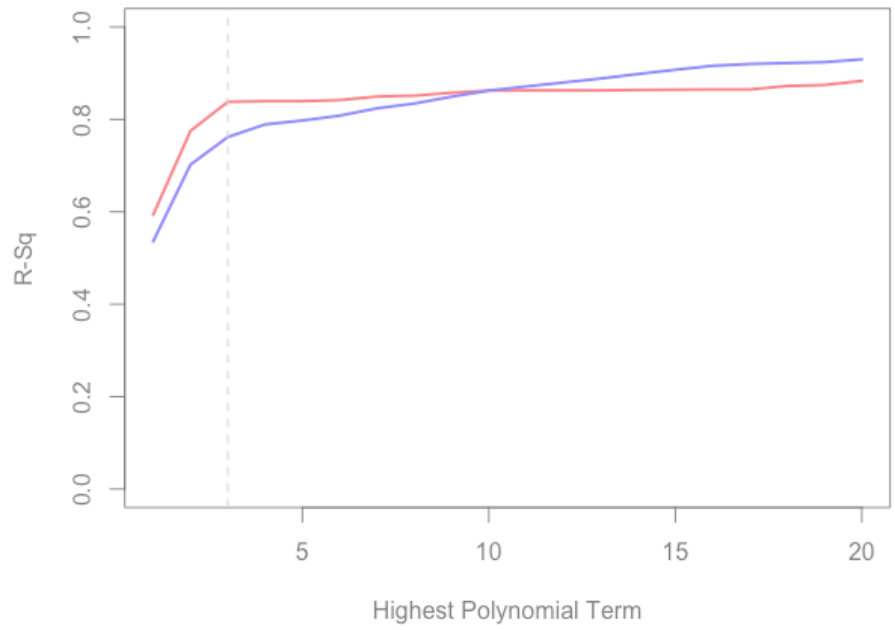
Question 11

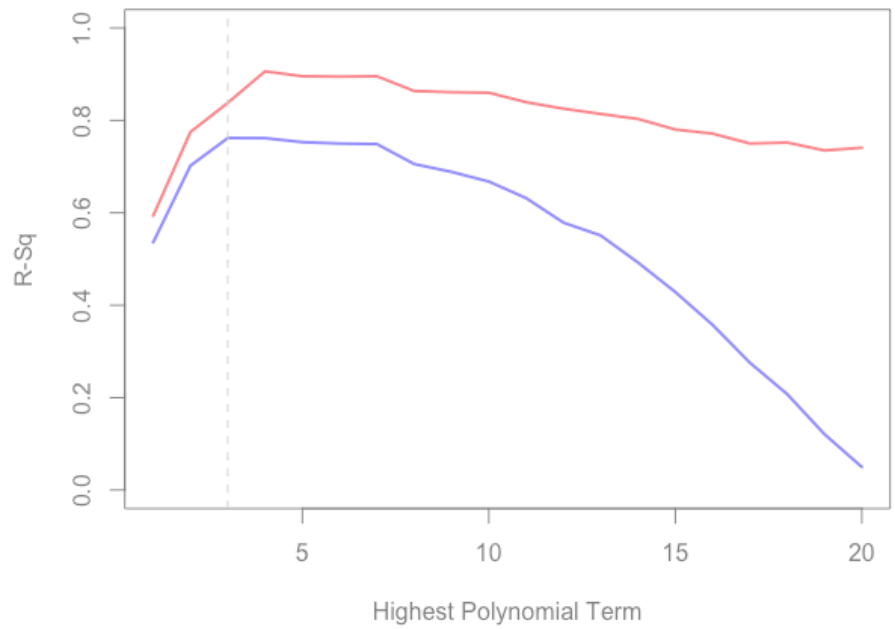
1 / 1 pts

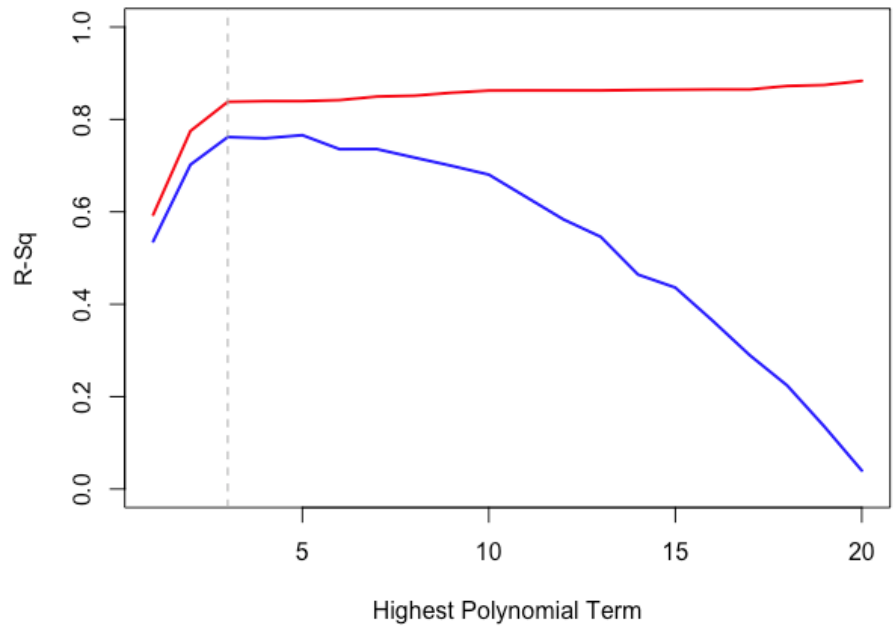
The order-1, order-2, and order-3 polynomial regression models were fit on a training set of 160 observations. The resulting R^2 were calculated for each of the 3 models on the training set (shown in red) and on a separate testing set of 40 observations (shown in blue). Determine which of the complete figures is the correct one for polynomial models up to order-20.




R-Sq
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Highest Polynomial Term


R-Sq
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Highest Polynomial Term


R-Sq
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Highest Polynomial Term


R-Sq
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20
Highest Polynomial Term

Question 12

1 / 1 pts

3 models were fit on the same set of data (3 quantitative predictors were used):
1) ordinary least squares model (OLS)
2) LASSO
3) Ridge regression

Match the resulting equations below to the approach that created them.
(A) $\hat{y} = -0.474 + 0.552x_1 + 0x_2 + 0.958x_3$
(B) $\hat{y} = -0.719 + 1.144x_1 + 0.215x_2 + 1.027x_3$
(C) $\hat{y} = -0.705 + 0.881x_1 + 0.154x_2 + 0.997x_3$

OLS B

LASSO A

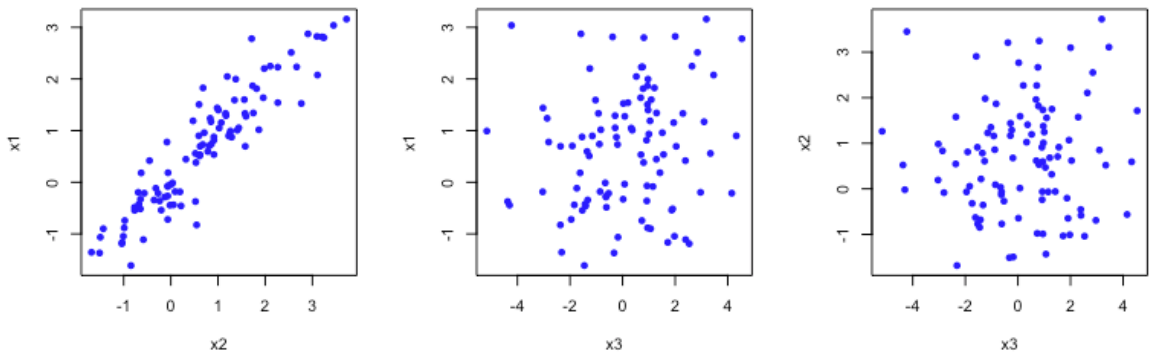
RIDGE C

Incorrect

Question 13

0 / 1 pts

3 predictors (x1, x2, x3) were considered for linear regression to predict y. Below are the scatterplots for x1 vs. x2, x1 vs. x3, and x2 vs. x3 (going left to right), along with the 3 separate simple linear regression models.



$\hat{y} = 0.19 + 1.28x_1$

$\hat{y} = 0.33 + 1.01x_2$

$\hat{y} = 0.90 + 0.87x_3$

Choose the most reasonable choice for the actual multiple regression model using all 3 predictors.

- ◉ $\hat{y} = 1.42 + 1.28x_1 + 1.01x_2 + 0.87x_3$
- ○ $\hat{y} = 0.22 + 1.07x_1 - 0.02x_2 + 0.75x_3$
- ○ $\hat{y} = 0.52 - 1.07x_1 - 0.89x_2 + 1.72x_3$
- ○ $\hat{y} = 0.07 + 1.37x_1 + 1.04x_2 + 0.03x_3$

Incorrect

Question 14

0 / 1 pts

Based on the previous question's data set (where the 3 scatterplots and 3 simple regression models are given), an analyst decides to perform a Principal Components Analysis (PCA). Choose the most reasonable set of $(\phi_{11}, \phi_{21}, \phi_{31})$, the coefficients that determine the first PCA predictor Z_1.

- ○ (1.28, 1.01, 0.87)
- ○ (0, 1, 0)
- ○ (0.33, 0.33, 0.34)
- ○ (0.70, 0.69, 0.18)
- ◉ (0.27, 0.26, 0.47)

It is Oct 13, 2016. NASA's radars discovered a small, 3 meter, iron base meteorite that just entered the Earth's atmosphere. A small meteorite will not create a wide-spread devastation but will still be dangerous for the citizens. Local authorities would like to know the location of the impact point so they that can warn residents and allocate resources based on the population affected.

The Governor has sought out the best data scientist in the state - you - to help save the day!

You are given two data sets:

1. Radar position estimates (x, y, z - coordinates; z being the altitude) of the meteorite at various times are available here (https://cs109alabs.github.io/lab_files/). x, y, z coordinates are in kilometers and time is in seconds.

2. Locations and other details of every dwelling in the town are provided **here** .

I. Using methods you learned in class to estimate the expected point of impact along with the region with

90% certainty.

II. Using the dwelling database, estimate the total number of people that will most likely be affected within this region.

AC209 students only: Additional measurements from another radar are available **here**. The accuracy of this radar is approximately 5 times higher than that of the first radar. Your model should take into account both radar data sets.

Question 15

Not yet graded / 1 pts

Enter your best estimate of the impact x,y-coordinates.

Your Answer:

impact estimate for x = 7213.09625879
impact estimate for y = 6806.57560265

Incorrect

Question 16

0 / 1 pts

Enter your best estimate of the number of people in the 90% certain impact region.

1994.0000

Question 17

Not yet graded / 36 pts

Upload your IPython notebook (the .ipynb file only).

Farmer_Rick_CS109A_Midterm.ipynb (https://instructure-uploads.s3.amazonaws.com/account_18750000000000001/attachments/2987668/Farmer_Rick_CS109A_Midterm.ipynb?AWSAccessKeyId=AKIAJFNFXH2V2O7RPCAA&Expires=1476537815&Signature=YVo0NOi6wX%2FeD7XsY8qiMufB2iQ%3D)

Quiz Score: **7** out of 52