**ECE 580 Homework #2**

Libo Zhang (lz200)

***Cross Validation Section***

**Question 1 (a) Solution:**

Sub-question (i) – The equation of my proposed model #1 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

Logo

Description automatically generated

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of for this model is shown below.

Sub-question (v) – For this model, the value is very small while the value is very large. Therefore, I think my proposed model #1 appears to have greater consistency and greater systematic error across data sets.

**Question 1 (b) Solution:**

Sub-question (i) – The equation of my proposed model #2 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

Logo

Description automatically generated

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of for this model is shown below.

Sub-question (v) – For this model, the value is still small while the value is still large. Therefore, I think my proposed model #2 appears to have greater consistency and greater systematic error across data sets. However, compared with model #1, I think model #2 does not have such strong extent of the so-called “greater consistency and systematic error”, because the difference between bias and variance is smaller, or the distance between bias and variance is closer, which means the model is more balanced in terms of consistency, variability, and systematic error across data sets.

**Question 1 (c) Solution:**

Sub-question (i) – The equation of my proposed model #3 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

A picture containing logo

Description automatically generated

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of for this model is shown below.

Sub-question (v) – For this model, the value is relatively small while the value is relatively large. Although this still indicates that my proposed model #3 appears to have greater consistency and greater systematic error across data sets, I think the so-called “greater” extent is weaker compared with model #1 and model #2, because for model #3 the and the values are closer to each other. Therefore, I think my proposed model #3 is the most balanced considering the bias-variance trade-off, in terms of model consistency, variability, and systematic error across data sets.

**Question 1 (d) Solution:**

A summary table for my three proposed models is shown below.

|  |  |  |  |
| --- | --- | --- | --- |
| Model Number |  |  |  |
| Model #1 | 1.39068 | 1.92 | 1.38876 |
| Model #2 | 1.33314 | 3.41 | 1.32973 |
| Model #3 | 1.07876 | 3.80 | 1.07496 |

According to the above table, the order of my 3 models from least complex (high bias, low variance) to most complex (low bias, high variance) can be concluded below (“<” denotes model complexity).

This conclusion also demonstrates my expectation when I initially designed the 3 models, as model #1 uses 1 feature, model #2 uses 2 features, and model #3 uses 3 features. Therefore, without guessing, actually, the model complexity ordering, or the bias-variance trade-off performance of my 3 models should have already been very explicit, and what I am doing here is to prove my expectation and make it more solid.

**Question 1 (e) Solution:**

I would select model #3 as my ideal model. This is because model #3 has the lowest value of E{MSE} and the lowest value of , which means that model #3 has the best regression performance and the least systematic error across data sets. Although model #3 has the highest value of , the variance value of model #3 does not increase significantly compared to model #1 and model #2. According to bias-variance trade-off, it is very reasonable for me to select model #3 as my ideal model.

Here, my conclusion of selecting model #3 is similar to the conclusion I came to without cross-validation in Homework #1. In Homework #1, I also selected model #3 as my ideal model because model #3 had the highest unadjusted score and the highest adjusted score, which could tell me that all 3 features in model #3 are meaningful (can provide additional information), and model #3 has the best regression performance.

**Question 1 (f) Solution:**

Before formally submitting the complete print-out of my code, I want to show that I indeed structure and organize my code according to the requirements to make it modular and extensible.

I write a function named “model\_summary” integrating all of my previous self-defined functions. This function could generate all of the regression results in Homework 2 using a single code block that accepts:

1. A given data set (X and y for my regression model)
2. The desired norm
3. The desired regularization
4. The desired cross-validation parameters

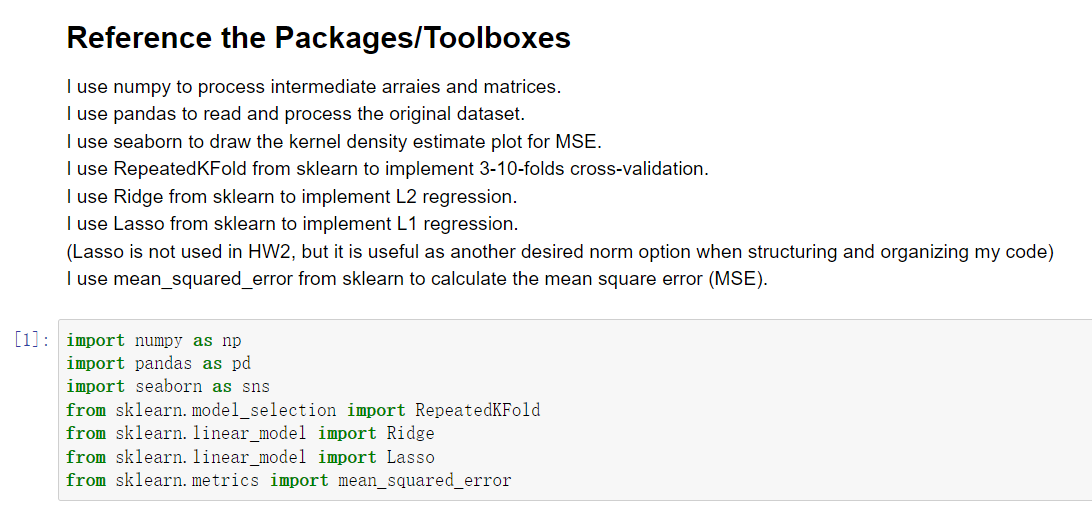
An example of “a single code block” is shown below.

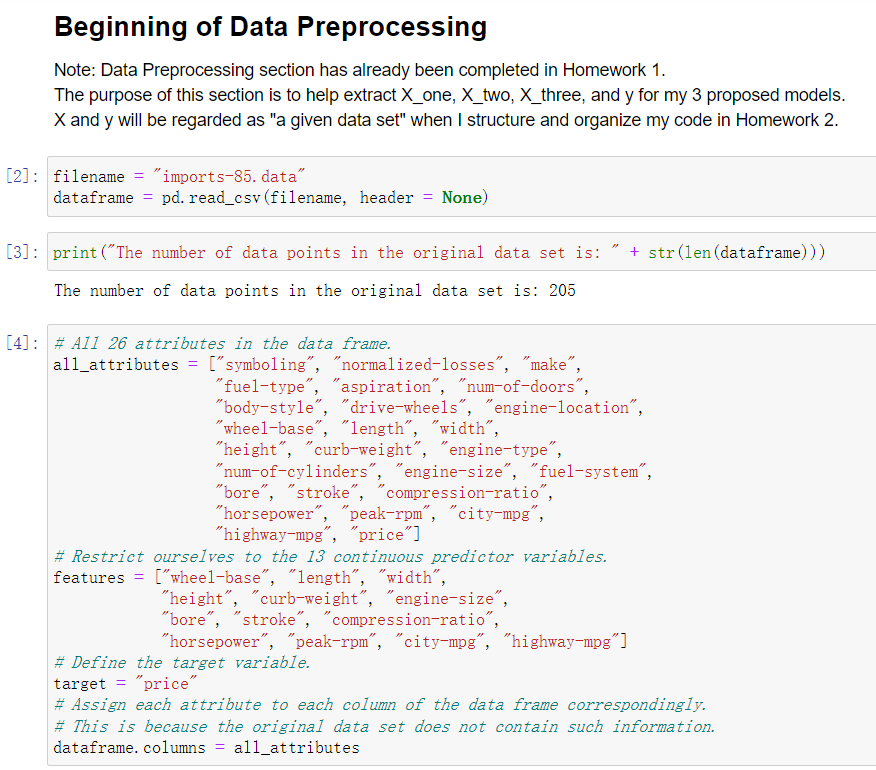
Chart, histogram

Description automatically generated

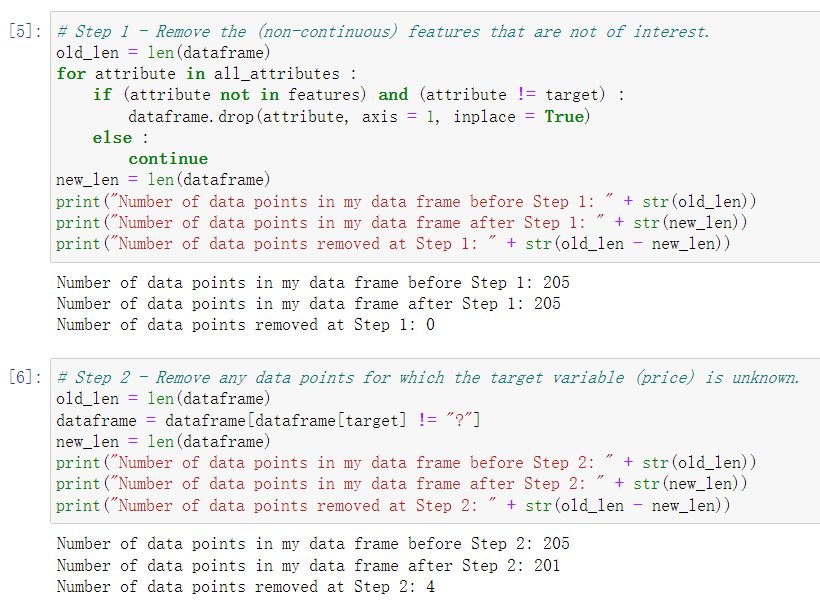
Please continue to the next page for the complete print-out of my code.

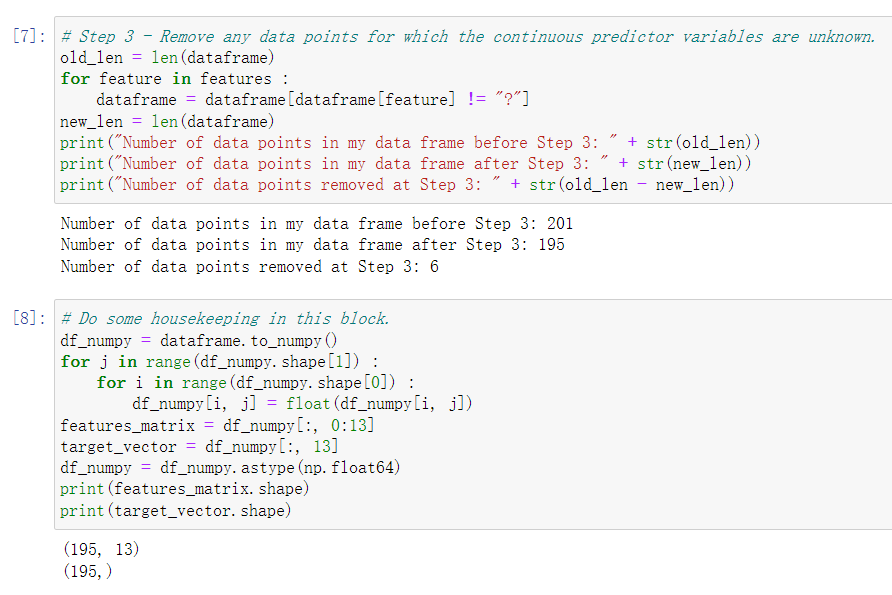
**Question 1 (f):**



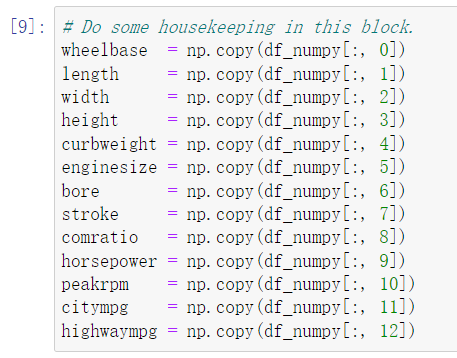


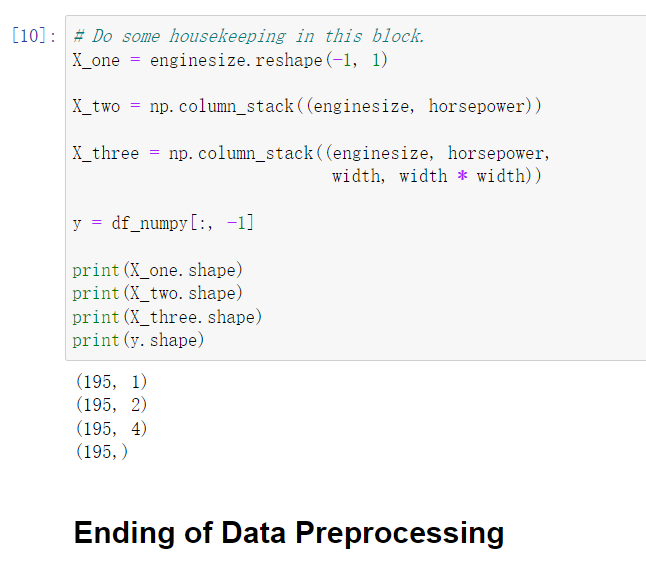
**Question 1 (f):**





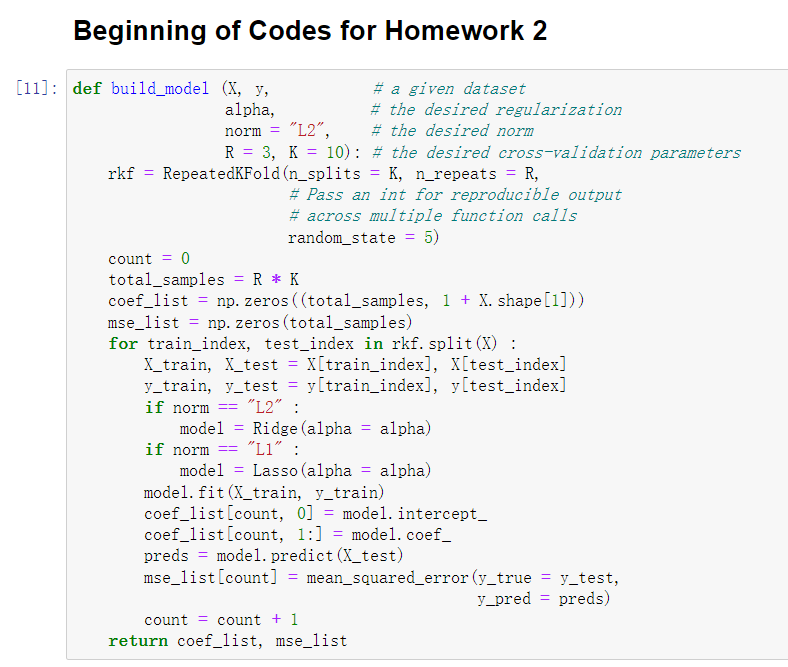
**Question 1 (f):**

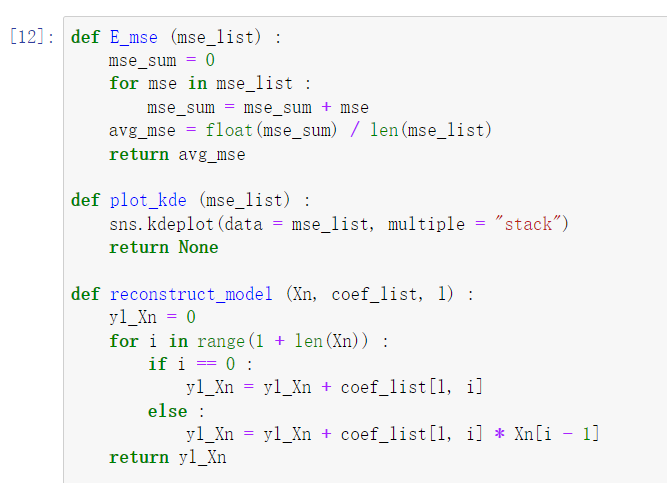




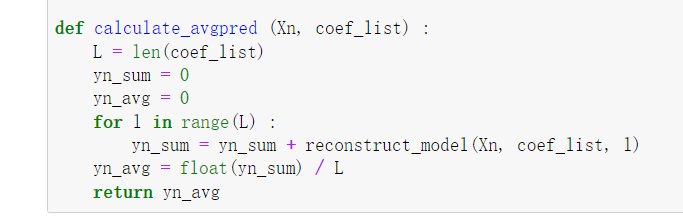
**Please continue to the next page for Homework 2 codes.**

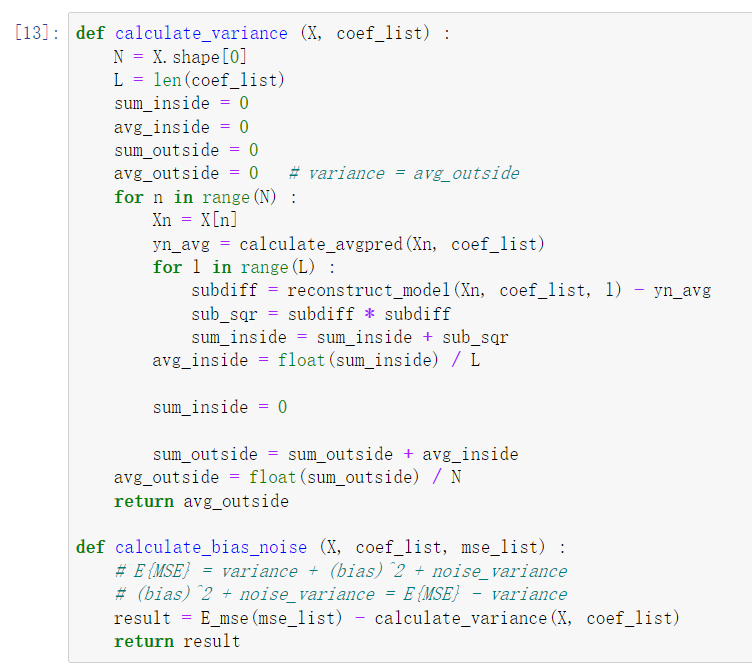
**Question 1 (f):**

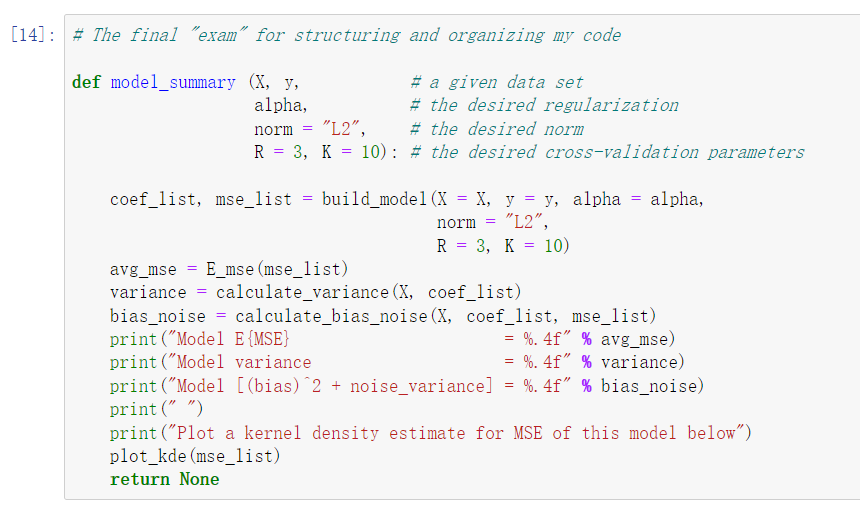




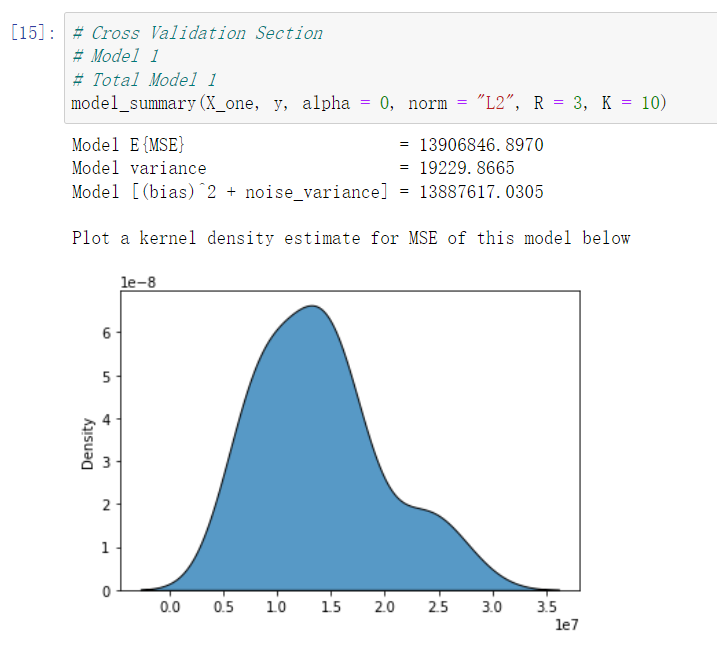
**Question 1 (f):**

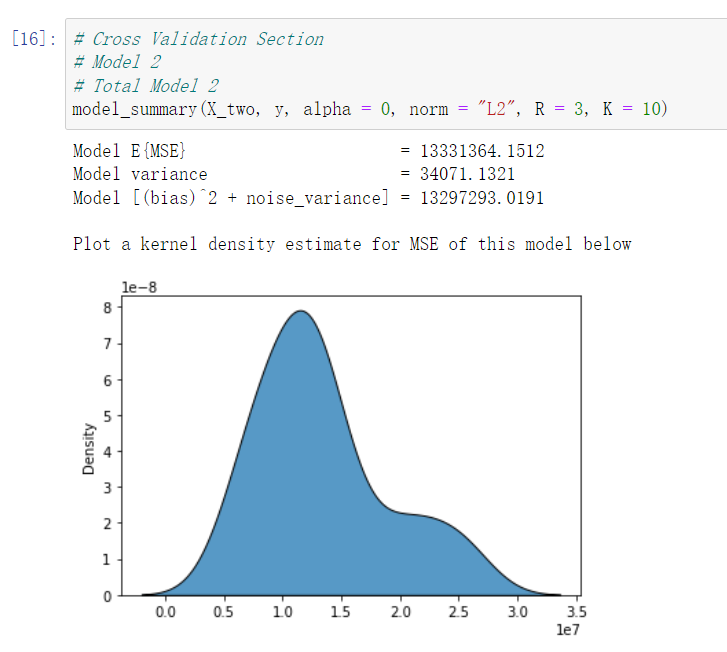




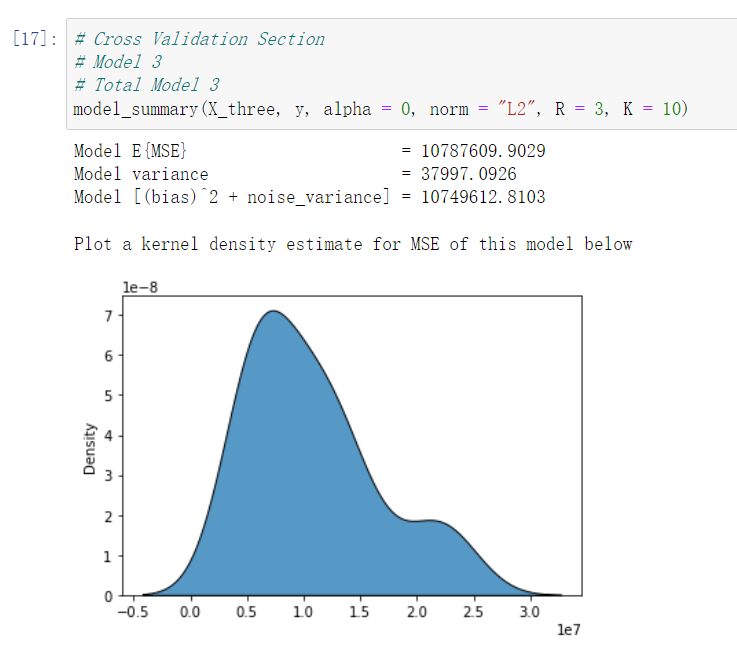


**Question 1 (f):**





**Question 1 (f):**



***Regularization Section***

**Question 2 (a) Solution:**

Note: When choosing a weighting for the regularization term (denoted as or alpha), my experiments show that a small alpha value does not affect my model #1 much, in terms of E{MSE}, variance, and bias. However, if the alpha value is too large, then my model #1’s regression performance will degrade significantly. Therefore, considering such trade-off, I choose for my proposed model #1.

Sub-question (i) – The equation of my proposed model #1 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

A picture containing logo

Description automatically generated

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of is shown below.

Sub-question (v) – For this model, the value is very small while the value is very large. Therefore, I think this model appears to have greater consistency and greater systematic error across data sets.

**Question 2 (b) Solution:**

Note: When choosing a weighting for the regularization term (denoted as or alpha), my experiments show that a small alpha value does not affect my model #2 much, in terms of E{MSE}, variance, and bias. However, if the alpha value is too large, then my model #2’s regression performance will degrade significantly. Therefore, considering such trade-off, I choose for my proposed model #2.

Sub-question (i) – The equation of my proposed model #2 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

Logo

Description automatically generated

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of for this model is shown below.

Sub-question (v) – For this model, the value is still very small while the value is still very large. Therefore, I think this model appears to have greater consistency and greater systematic error across data sets.

**Question 2 (c) Solution:**

Note: When choosing a weighting for the regularization term (denoted as α or alpha), my experiments show that a small alpha value could clearly affect my proposed model #3, such as increasing E{MSE} and bias, while decreasing variance. If the alpha value is too large, then my model #3’s regression performance will degrade significantly. Therefore, I choose for my proposed model #3.

Sub-question (i) – The equation of my proposed model #3 is shown below.

Sub-question (ii) – The kernel density estimate plot for the mean square error (MSE) is shown below.

Logo

Description automatically generated with medium confidence

Given the kernel density estimate plot for MSE and the value of E{MSE}, I believe E{MSE} is a representative summary statistic (a good approximation) for MSE, because the plot displays a trend of uniform distribution, and the value of E{MSE} is very close to the mean value of this uniform distribution.

Sub-question (iii) – The value of for this model is shown below.

Sub-question (iv) – The value of for this model is shown below.

Sub-question (v) – For this model, the value is relatively small while the value is relatively large. Although this still indicates that my proposed model #3 appears to have greater consistency and greater systematic error across data sets, I think such “greater” extent is weaker compared with my model #1 and model #2, because for model #3 the variance and bias values are closer to each other. Therefore, I think model #3 is still the most balanced considering the bias-variance trade-off.

**Question 2 (d) Solution:**

A summary table for my 3 proposed models with/without regularization is shown below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model  (# / Total #) | Regularization  () |  |  |  |
| #1 (#1) | 0 | 1.39068 | 1.92 | 1.38876 |
| #1 (#4) | 10000 | 1.39325 | 1.74 | 1.39151 |
| #2 (#2) | 0 | 1.33314 | 3.41 | 1.32973 |
| #2 (#5) | 10000 | 1.33450 | 3.00 | 1.33150 |
| #3 (#3) | 0 | 1.07876 | 3.80 | 1.07496 |
| #3 (#6) | 1 | 1.14510 | 3.61 | 1.14149 |

According to the above table, the order of my 3 models (considering both L2 regularization and cross-validation) from least complex (high bias, low variance) to most complex (low bias, high variance) can be concluded below (“<” denotes model complexity).

According to the above table, we can see that after adding the appropriate L2 regularization term, the E{MSE} tends to increase a little bit, the variance tends to decrease a bit, and the bias tends to increase a bit, but all changes are not significant. Therefore, we can also conclude that the results using L2 regularization (Ridge regression) are similar to the results using no regularization.

**Question 2 (e) Solution:**

I would select the proposed model #3 under no regularization as my ideal model. This model is also #3 among the Total # of models listed in the above summary table.

As for the reason, firstly, this model has the lowest E{MSE} and the lowest bias among all 6 models, which means that this model has the best regression performance and the least systematic error across data sets.

Secondly, although this model has the highest variance, its variance is not significantly higher than the other 5 models, and in fact, this model has the smallest difference (or the closest distance) between bias and variance, which means this model is the most balanced one among all 6 models, according to bias-variance trade-off.

Thirdly, during my experiments with Ridge regression, I find that neither a small L2 regularization term nor a large L2 regularization term could help my 3 models improve regression performance. Using Ridge regression in my models would always tend to increase E{MSE} and to increase the difference between bias and variance, making my models less balanced according to bias-variance trade-off.

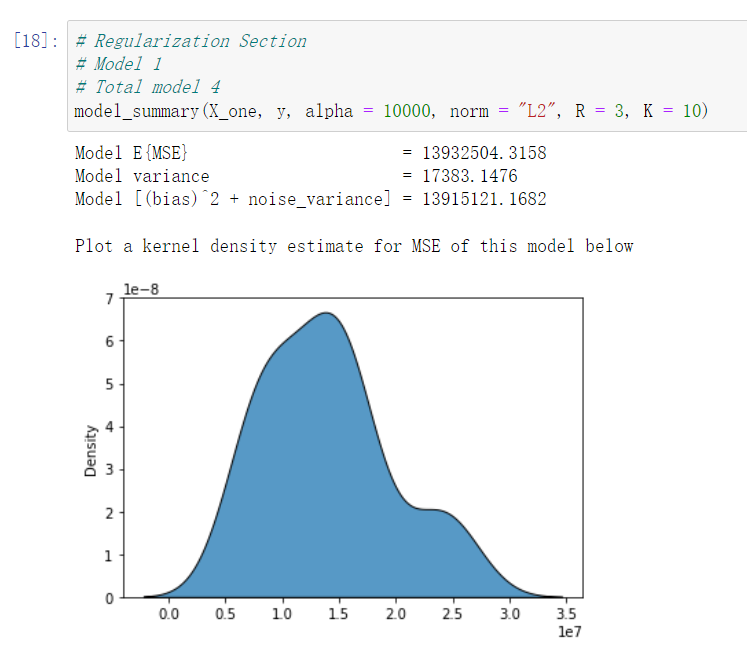
**Question 2 (f) Solution:**

Before formally submitting the complete print-out of my code, I want to show that I indeed structure and organize my code according to the requirements to make it modular and extensible.

I write a function named “model\_summary” integrating all of my previous self-defined functions. This function could generate all of the regression results in Homework 2 using a single code block that accepts:

1. A given data set (X and y for my regression model)
2. The desired norm
3. The desired regularization
4. The desired cross-validation parameters

An example of “a single code block” is shown below.



Please continue to the next page for the complete print-out of my code.

**Question 2 (f):**

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

**Question 2 (f):**

Text

Description automatically generated

Text

Description automatically generated

**Question 2 (f):**

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Please continue to the next page for Homework 2 codes.**

**Question 2 (f):**

Text

Description automatically generated

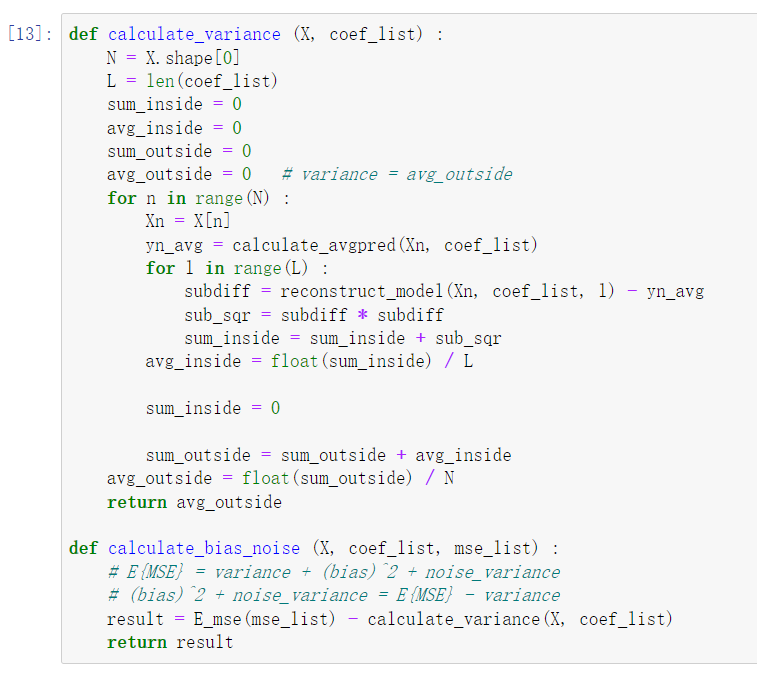
Text

Description automatically generated

**Question 2 (f):**

Graphical user interface, text, application, Word

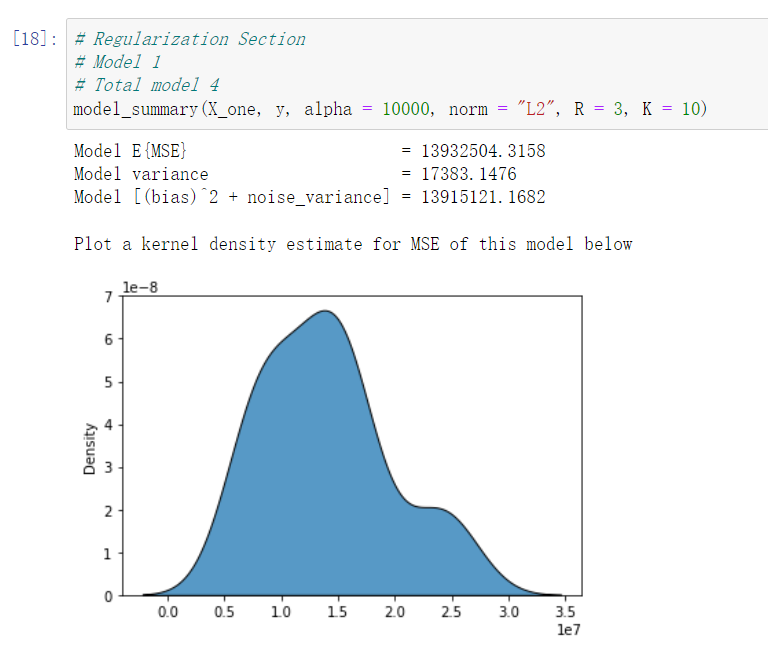
Description automatically generated

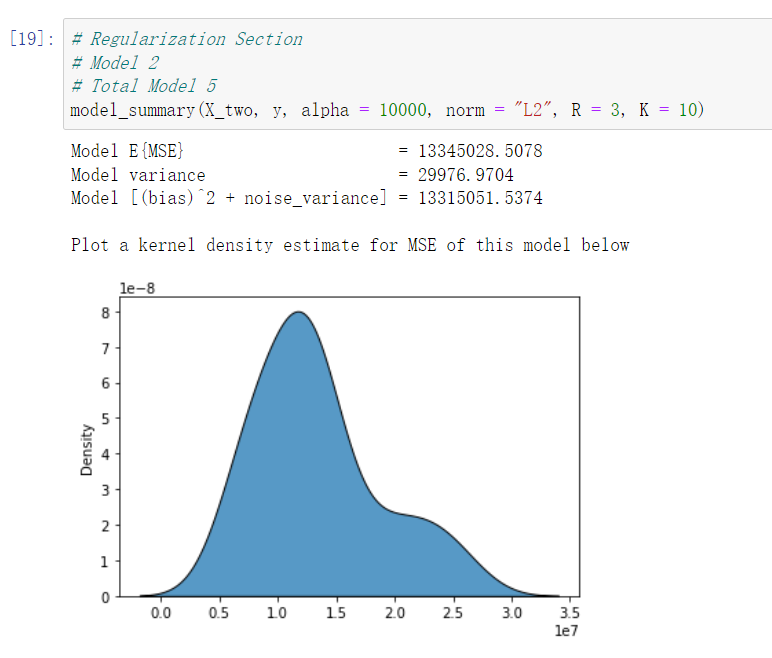


Text

Description automatically generated

**Question 2 (f):**





**Question 2 (f):**

