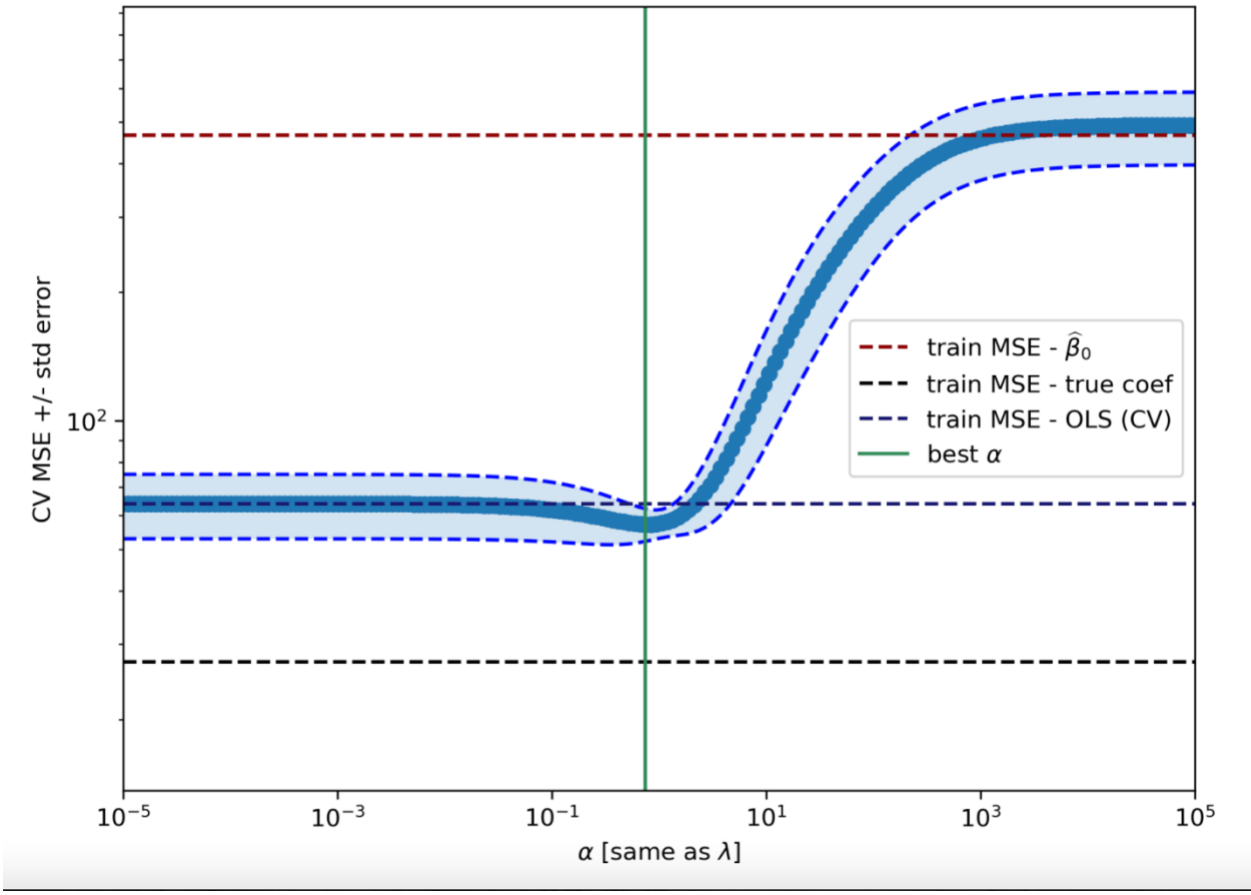
Michael Koopmann
CS 474
HW 4
10/21/21

1. Ridge Regression
    a. Model



    b. Brief Description
        i. What model is picked as λ → 0 is the OLS solution model (no penalty-- all features) because as lambda goes to zero it means that we're decreasing the penalty for regulation, so it'll pick the OLS model which has no penalty.
        ii. What model is picked as λ → ∞ is the just intercept model because as lambda goes to infinity the coefficients flatten out to zero and the β s will go to zero. So, the model will be pulled more and more towards the mean of the Y which is the intercept model.
    c. The largest coefficient was 9.17178095 and the smallest was 0.08404492
    d. Model Picked by Ridge Regression
        i. Ridge picked 40 features.
        ii. True Model picked 20 features.

iii. There were 40 features in the data set.

e. The results for d above were different for the number of features used between ridge and lasso with ridge using 40 and lasso using 25 this is because lasso performs feature selection in which it selects the best features will turning others to absolute zero so basically not selecting them while ridge used all the features.

2. Subset Searches

a. Models

i. Best Subset Selection

```
print('\n %2i  %.2f  %s'%(k,mse_train,str(subset)))
```

```
 0  168.68  ()
 1  116.71  (5,)
 2  71.93  (3, 5)
 3  46.66  (3, 5, 10)
 4  35.02  (3, 5, 6, 10)
 5  27.17  (3, 5, 6, 7, 10)
 6  26.72  (3, 5, 6, 7, 9, 10)
 7  25.88  (3, 5, 6, 7, 8, 9, 10)
 8  25.73  (1, 3, 5, 6, 7, 8, 9, 10)
 9  25.68  (1, 2, 3, 5, 6, 7, 8, 9, 10)
10  25.68  (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
```

ii. Forward Selection

```
print('\n %2i  %.2f  %s'%(k,mse_train,str(fsubset)))
```

```
 0  168.68  ()
 1  116.71  (5,)
 2  71.93  (3, 5)
 3  46.66  (3, 5, 10)
 4  35.02  (3, 5, 6, 10)
 5  27.17  (3, 5, 6, 7, 10)
 6  26.72  (3, 5, 6, 7, 9, 10)
 7  25.88  (3, 5, 6, 7, 8, 9, 10)
 8  25.73  (1, 3, 5, 6, 7, 8, 9, 10)
 9  25.68  (1, 2, 3, 5, 6, 7, 8, 9, 10)
10  25.68  (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
```

```
[244]: for k in range(p+1):
```

iii. Backward Selection

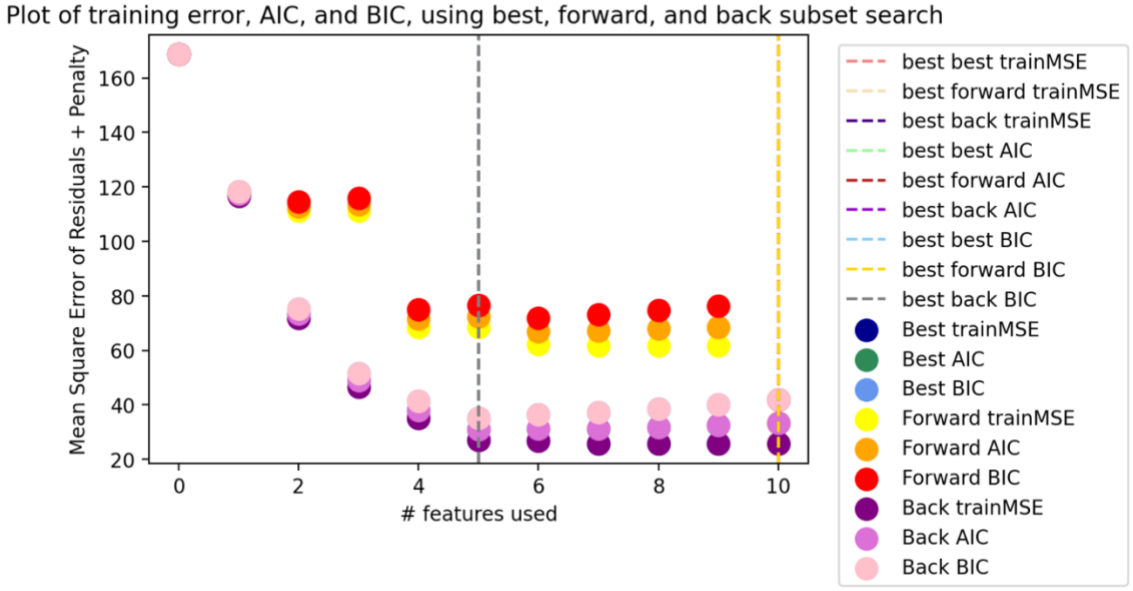
```
print('\n %2i  %.2f  %s'%(k,mse_train,str(bsubset)))
```

```
 0  168.68  ()
 1  116.71  [5]
 2  71.93  [3, 5]
 3  46.66  [3, 5, 10]
 4  35.02  [3, 5, 6, 10]
 5  27.17  [3, 5, 6, 7, 10]
 6  26.72  [3, 5, 6, 7, 9, 10]
 7  25.88  [3, 5, 6, 7, 8, 9, 10]
 8  25.73  [1, 3, 5, 6, 7, 8, 9, 10]
 9  25.68  [1, 2, 3, 5, 6, 7, 8, 9, 10]
10  25.68  [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```

1. All my selections had the same sets and MSE scores for each layer.

b. Plot

```
plt.legend(loc='best', bbox_to_anchor=(1, 0.5, 0.5, 0.5))
# plt.xlim(-0.5, p+10)
plt.show()
```



i.

c. Each set selected 5

d. Each set selected 10 I assume that they're the same because the models they produced were the same as each other as well for the data set.