Regression

1) The estimated weights and biases were very close to the real values. The mean error in the 20 estimated weights was equal to 0.01921, and the error in the estimated bias was equal to 0.00828. The significance of the features was mostly indicative of the underlying nature of the data. The most significant features were $X_1, \ldots, X_5$ with weights of the order of 10^{-1} and the least significant features were X_{16}, X_{17}, X_{19}, X_{20} with weights of the order of 10^{-3}. However, no feature was pruned. The true error of the model when tested on a dataset of 1,000,000 samples was equal to 0.1027.
2) The true error of the ridge regression model as a function of λ is shown in Fig 1.

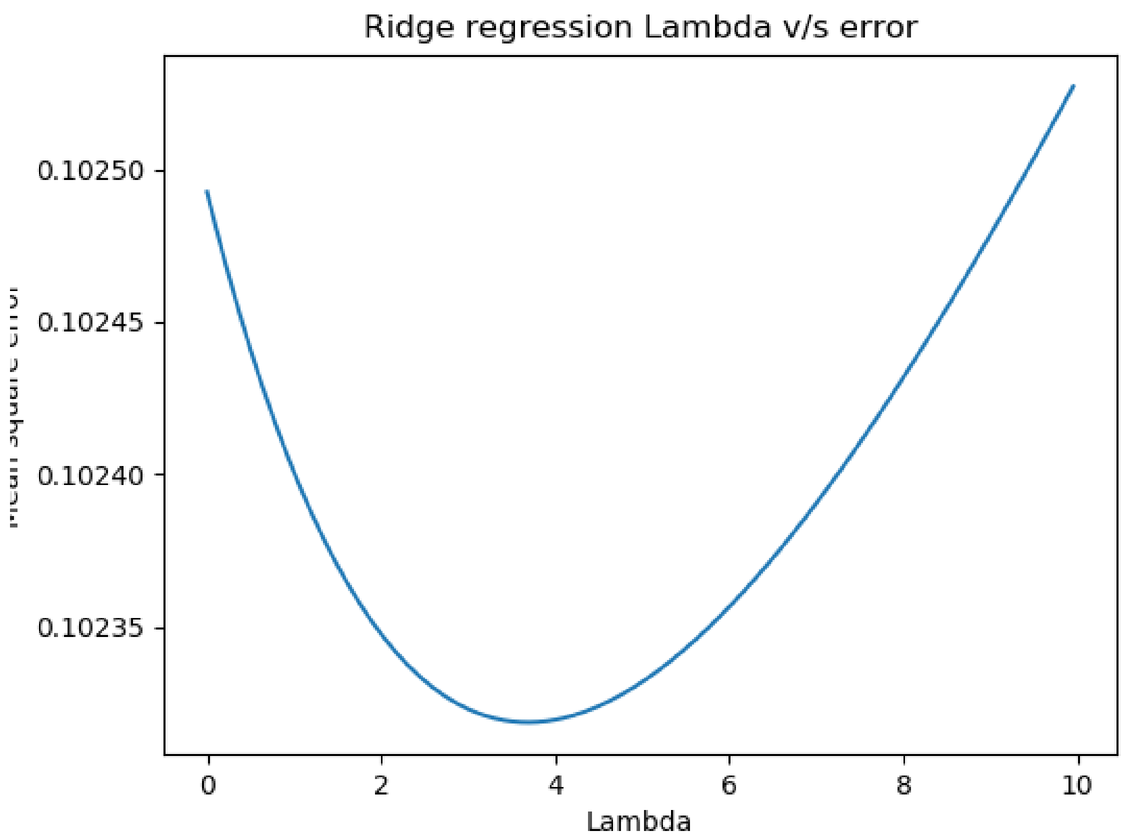

Fig 1. Ridge regression – True error v/s λ

The optimal λ was obtained as 3.7. The weights obtained at this λ are equal to

[5.84183645e-01]
[2.92793489e-01]
[2.17721692e-01]
[1.50465595e-01]
[1.05246998e-01]

[6.29385534e-02]
[3.21919604e-02]
[1.80273016e-02]
[2.06731509e-02]
[1.79502276e-02]
[7.53708286e-03]
[1.25943215e-02]
[-5.24324523e-02]
[1.17825541e-02]
[3.50514611e-02]
[-9.40122393e-03]
[-1.82180084e-03]
[2.98515146e-02]
[-4.52842424e-03]
[-1.71810246e-03]

The estimated bias was 10 -1.71737624e-16. The most and least significant features stayed mostly the same as those in the case of naïve least squares method. As in the case of naïve least squares, no features were pruned. A comparison of some basic metrics between the naïve least squares and ridge models is presented in Fig 2. The ridge model is the clear winner.

Metric	**Naïve least squares**	**Ridge regression**
Mean error in weights	0.019211	**0.0172013**
Var in error in weights	0.000348	**0.0000295**
Mean error in bias	0.00828	**0.00828 (Lower at 17th decimal)**
True model error	0.102699	**0.1025003**

Fig 2. Naïve least squares v/s Ridge regression

3) The number of pruned features as a function of λ is shown in Fig 3. Clearly, the number of pruned features increases as λ increases, until only the bias in non-zero.

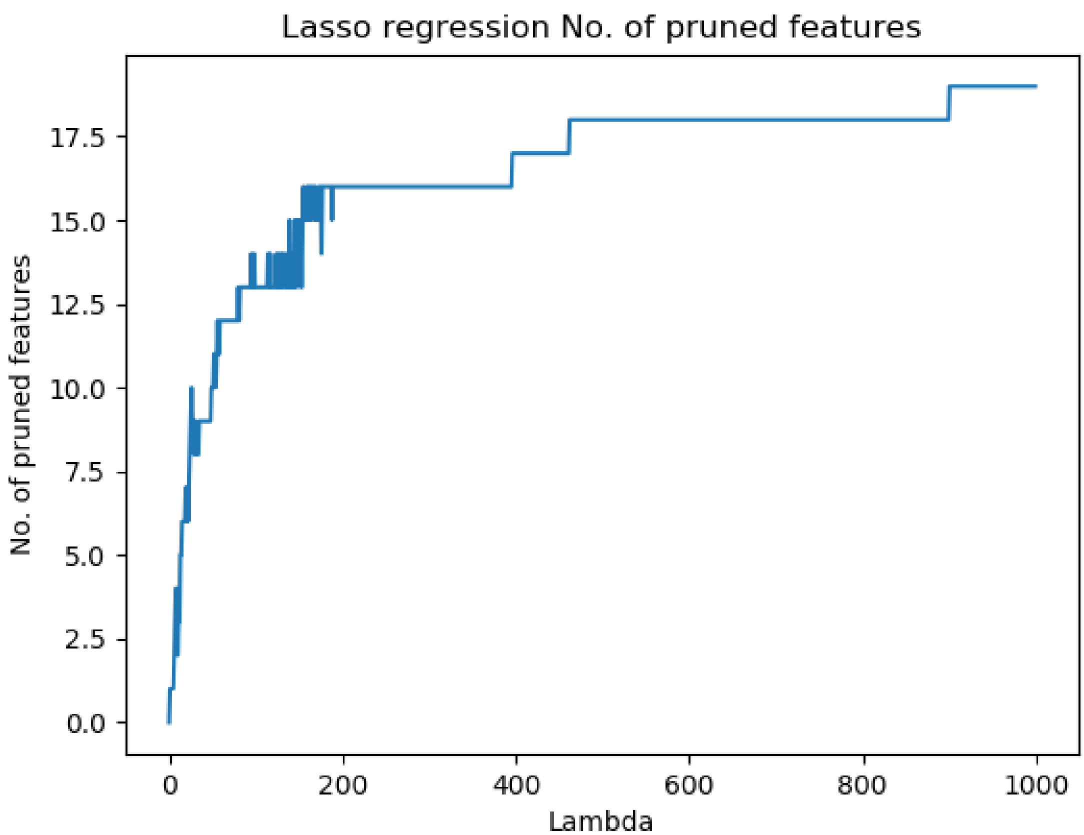

Fig 3. Number of pruned features v/s λ

4) The true error the lasso regression model as a function of λ is shown in Fig 4.

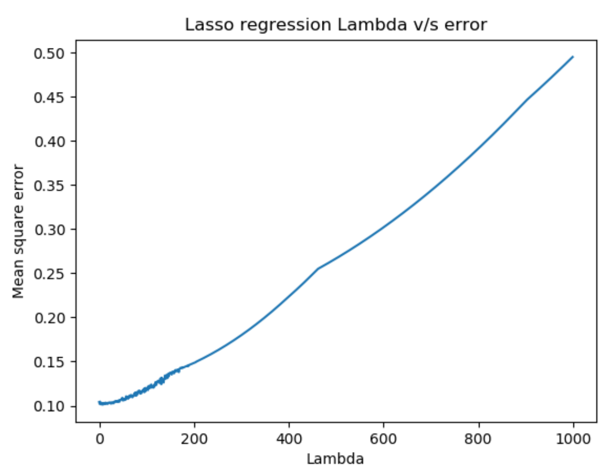

Fig 4. True error v/s λ

The optimal value of λ was obtained as 4. The weights obtained at this λ were

[5.90000085e-01]
[3.32405739e-01]
[1.95784499e-01]
[9.10458651e-02]
[6.71651124e-02]
[6.06602840e-02]
[3.05992293e-02]
[1.53458694e-02]
[1.88962719e-02]
[1.62889020e-02]
[0.00000000e+00]
[4.14759679e-02]
[0.00000000e+00]
[9.67363552e-03]
[0.00000000e+00]
[-7.23479261e-03]
[-2.87599178e-06]
[2.72409780e-02]
[-2.61272583e-03]
[0.00000000e+00]

The estimated bias was 10-1.52766688e-16. The most significant features were X_1, X_2, X_3 with weights of magnitude 10^{-1}. The least significant were X_{11}, X_{13}, X_{15}, X_{20} which were successfully pruned w/ weight 0. The comparison between naïve least squares and lasso regression is presented in Fig 5.

Metric	**Naïve least squares**	**Ridge regression**
Mean error in weights	0.019211	**0.01161**
Var in error in weights	0.000348	**0.00015**
Mean error in bias	0.00828	**0.00828 (Lower at 17th decimal)**
True model error	0.102699	**0.10210**

Fig 5. Naïve least squares v/s Lasso regression

5) As mentioned in Q4, the relevant features at the optimal λ with respect to mean square error of the model were all features except X_{11}, X_{13}, X_{15}, X_2. Only the set of relevant features were estimated against using ridge regression with regularization parameter λ

= 4. This value of λ is the same as the optimal λ obtained for Lasso regression. While this may not be optimal, it is still a good value as it is expected to increase the weightage of the L2 norm in the objective function as it did for the L1 norm without overpowering the error term. A comparison of the combined Lasso-ridge model v/s the naïve least squares on test data of 1,000,000 samples is presented in Fig 6.

Metric	**Naïve least squares**	**Lasso-Ridge**
Mean error in weights	0.019211	**0.008945**
Mean error in bias	0.008282	**0.008282 (Lower at 16th decimal)**
True model error	0.110287	**0.102483**

Fig 5. Naïve least squares v/s Lasso regression

SVM

1) One simple way of generating alphas is:

$$\alpha_i = \begin{cases} \dfrac{K}{no.of - ve\ samples} & if\ y_i\ is\ a - ve\ sample \\ \dfrac{K}{no.of + ve\ sample} & if\ y_i is\ a + ve\ sample \end{cases}$$

This handles all the constraints, and the multiplier K>0 ensures that α is away from the boundary i.e. α_i =0. We can make sure to not step outside the constraint region by decreasing ϵ_t by a small amount each iteration. The initial ϵ_t was selected as 1 and it was multiplied by a factor of x0.5. This makes sure, that the value of ϵ_t never crosses 0 and also that the difference between any 2 iteration steps is <=0.5.

2) With an initial value of $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 5$, the solver converged to the correct values of $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.125$ as obtained by hand after ~250 iteration.

3) The primal SVM can be reconstructed using the formula:

$$\acute{\omega} = \sum_i \alpha_i y_i x_i$$

$$b = y_i - \acute{\omega}\, x_i\ if\ \alpha_i = 0$$

On using the XOR data against the calculated values of α_i, the primal SVM was able to classify the XOR as expected.