

1 Linear Regression

1. As fig. 1 and 2 shows, the weights and bias of the model I got were pretty close to the true weights and bias. But there was still some noise in the weights of my model.

My model concluded x_1 as the most significant feature, and x_{19} as the least significant feature. Due to the insufficient of the data, my model concluded x_{19} as the least significant feature. When the size of data set becomes 100,000, the weights of my model are shown in fig. 3, which are indeed pretty close to the true weights.

When $m = 100,000$, the ‘true’ error I got was 0.10375875.

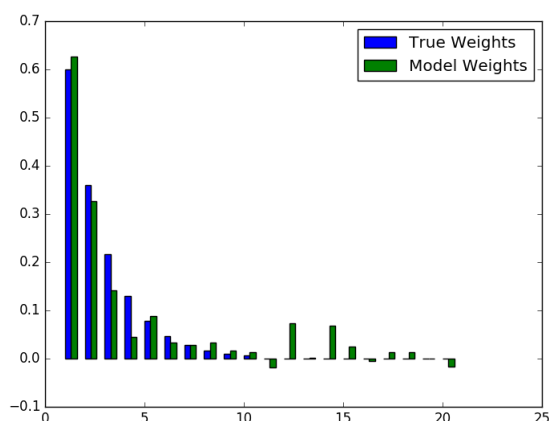


Figure 1: Comparison of true weights and the weights of model.

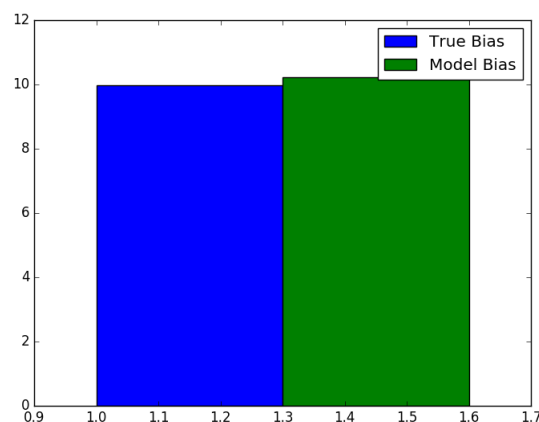


Figure 2: Comparison of true bias and the bias of model.

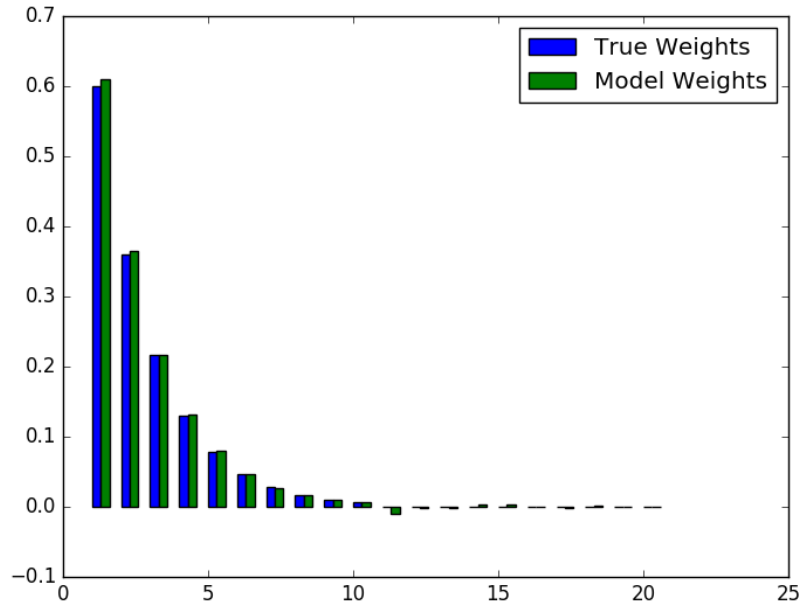


Figure 3: The weights of model when $m = 100,000$.

2. As fig. 4 shows, the optimal λ is 0.05, which yields an optimal error of 0.10415862.

As fig. 5 and 6 shows, the bias of the model I got was pretty close to the true bias. But the weights were not really ideal, because it should have pruned some features. There's still some space for improvement.

My model concluded x_1 as the most significant feature, and x_{18} as the least significant feature. Due to the insufficient of the data, my model concluded x_{18} as the least significant feature. When the size of data set becomes 100,000, the weights of my model are shown in fig. 7, which are indeed pretty close to the true weights.

When $m = 1000$, the model I got using ridge regression was somehow similar to that of naive linear regression, but have smaller weights.

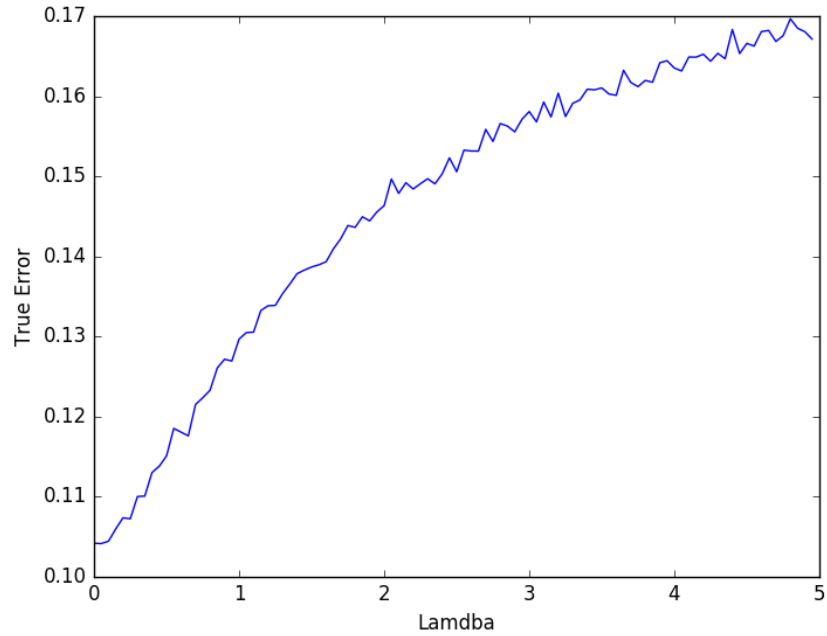


Figure 4: True errors vs lambdas.

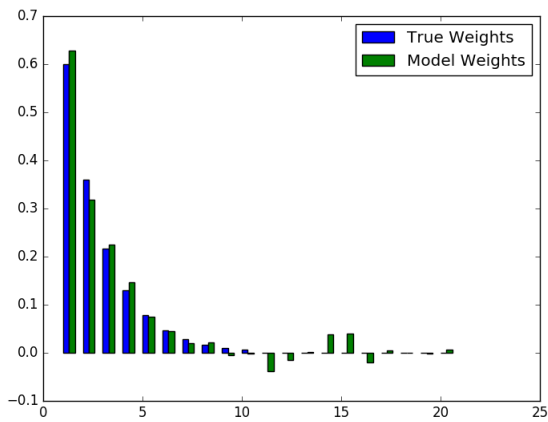


Figure 5: Comparison of true weights and the weights of model.

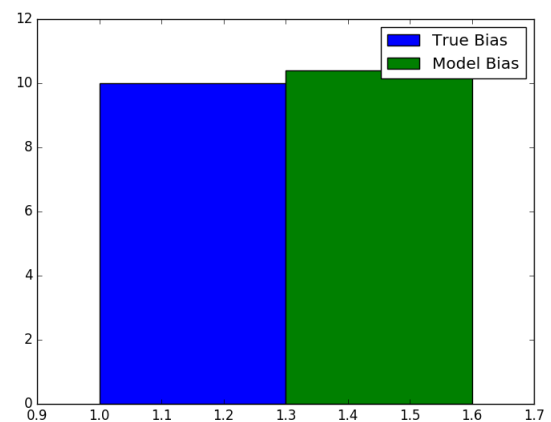


Figure 6: Comparison of true bias and the bias of model.

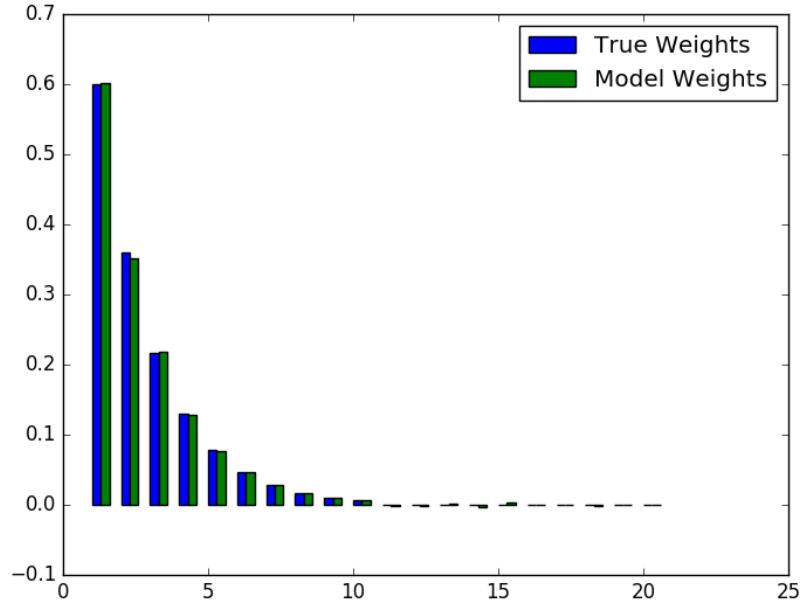


Figure 7: The weights of model when $m = 100,000$.

3. As fig. 8 shows, as λ increases, the number of zero-weighted features increases.

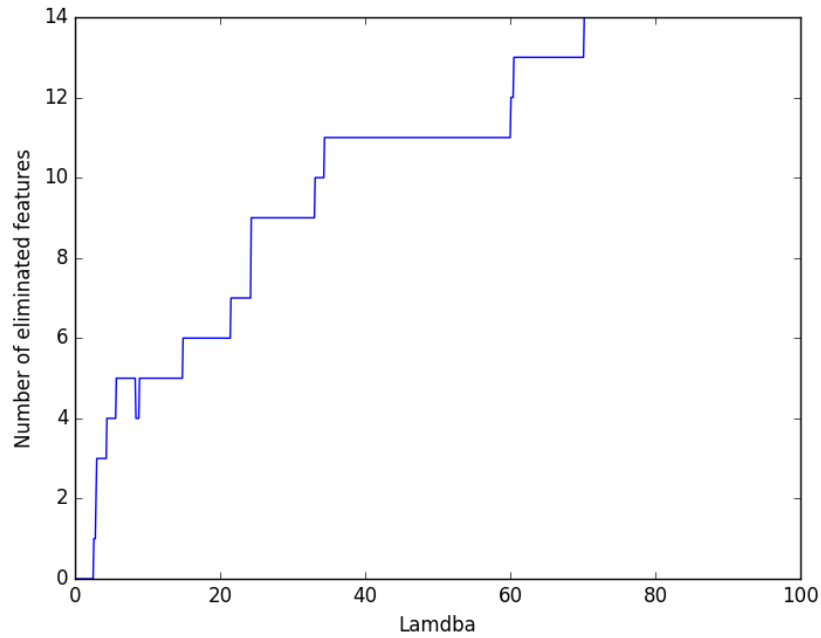


Figure 8: Number of eliminated features vs lambda when $m = 1000$.

4. As fig. 9 shows, the optimal λ is 17.4, which yields an optimal error of 0.09864528.

As fig. 10 and 11 shows, the bias of the model I got was pretty close to the true bias. Most of the weights were close to the true weights, especially weights for those unrelated features. But there are still some strange behaviors for $x_4, x_5, x_{11}, x_{12}, x_{13}$.

My model concluded x_1 as the most significant feature, and $x_4, x_{11}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}$ as the least significant features as their weights were 0.

When $m = 1000$, the model I got using lasso regression was actually worse than the naive model (In my opinion), despite the lasso model eliminated many features.

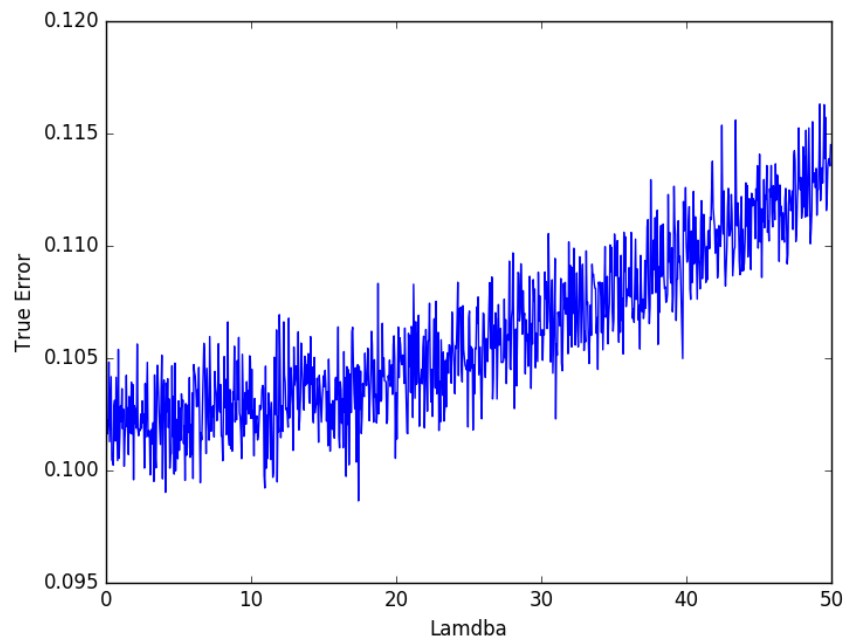


Figure 9: True errors vs lambdas.

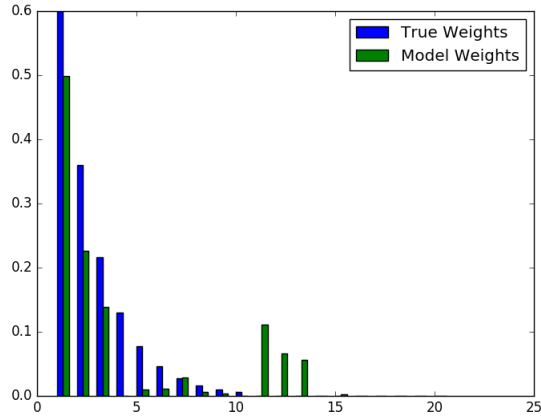


Figure 10: Comparison of true weights and the weights of model.

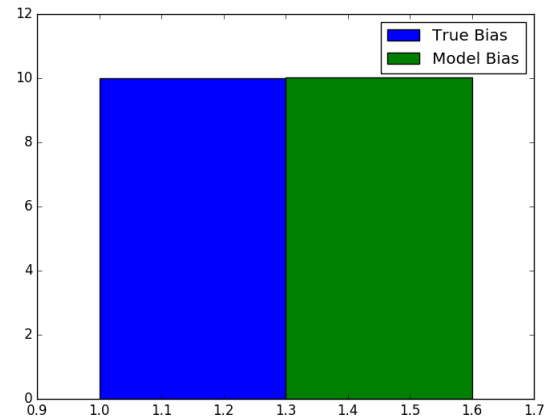


Figure 11: Comparison of true bias and the bias of model.

5. I also used brute-force to find the best ridge regression regularization constant. As fig 12. shows, the best constant I found was 0.05.

As fig. 13 and 14 shows, the weights of the model were not as close to the true weights as that of naive model did.

My model concluded x_1 as the most significant feature, but it also concluded that x_2, x_{15} were significant. The least significant features were $x_4, x_{11}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}$ as their weights were 0.

The testing error of lasso-ridge combination model was 0.10367449. Compared with the results of naive model, the lasso-ridge combination model was actually worse.

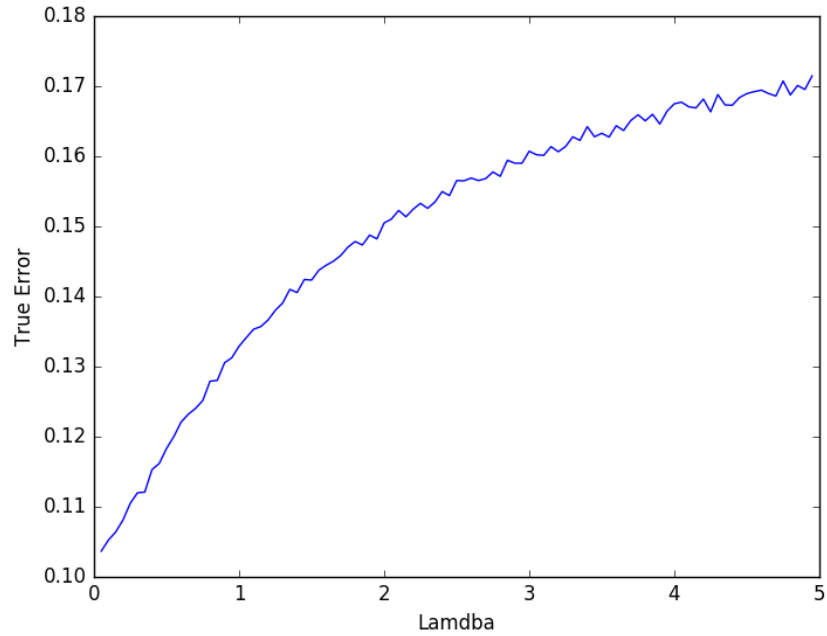


Figure 12: True errors vs lambdas.

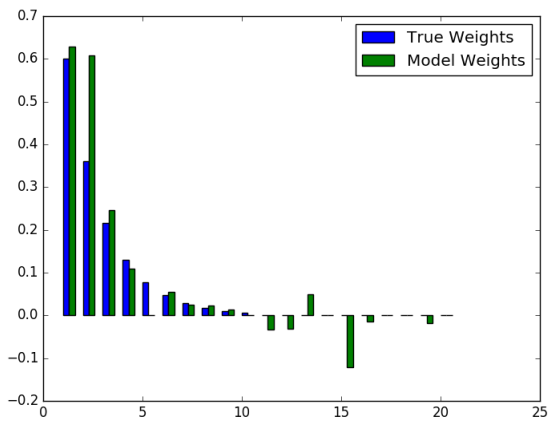


Figure 13: Comparison of true weights and the weights of model.

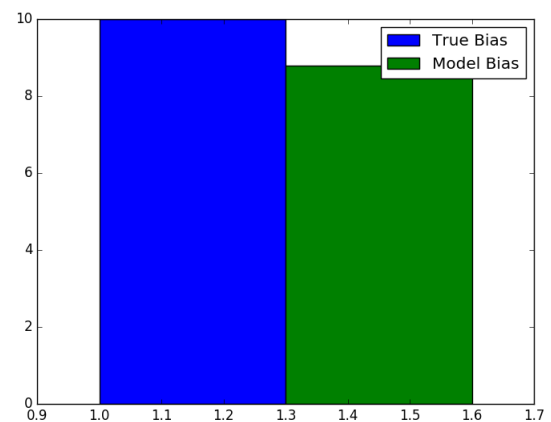


Figure 14: Comparison of true bias and the bias of model.