

Assignment 3

William Richard

October 11, 2012

1 Regularized Linear Regression

1.1 How does λ affect the results?

As we would expect, for the train data, the MSE approaches the expected value as we increase λ . On the test data, the MSE begins to approach the expected value, but at a very small λ hits an inflection point and begins to go away from the expected MSE value. This implies that, for larger λ , our model has overfit the data, and that the real model should have a small λ , i.e. a small regularization, and be trained mostly based on the actual trainin data.

1.2 How does the choice of λ depend on the setting of features or number of examples?

To explore how the number of examples affects the choice of λ , please compare figures 1.2 and 1.2. Figure 1.2 clearly does better with a smaller λ than figure 1.2. With fewer examples, the 150(1000)-100 case becomes dominated by λ , and thus behaves better when λ is smaller, or having a smaller effect on the output. Alternately, the 1000-100 case has many more examples, and thus can absorb the effects of a larger λ . The large λ has the added bonus of normalizing the data to a greater extent.

Next, compare figures 7 and 1.2 for the 100-100 and 100-10 cases, respectively. These two figures would imply that with fewer features, in the 100-10 case, a lower λ is better.

Figure 1: Regularized Linear Regression on 1000-100

Figure 2: Regularized Linear Regression on 100-10

2 Learning Curves

With the larger λ values, you see much faster stabilization of MSE value with respect to sample size. This occurs because, with larger λ , individual samples are normalized more, thus there is less noise, and the data matters less, resulting in faster convergence to the final MSE value.

3 Cross Validation

Data Set	Cross Validation		Reg. Lin. Re- gression	
	λ	MSE	λ	MSE
50(1000)-100	24	5.3	8	5.541
100(1000)-100	30	4.840	19	5.206
150(1000)-100	46	4.869	23	4.849
1000-100	39	4.137	27	4.316
100-10	12	4.159	8	4.160
100-100	20	4.494	22	5.078
wine	3	0.642	2	.0625

In most of the data sets, especially the ones with a large number of examples, the λ that was chosen using regularized linear regression was close to the λ chosen using cross validation. When cross validation gave a different result than regularized linear regression, cross validation has produced a vastly superior result.

Based on the 50(1000)-100, 100(1000)-100, 150(1000)-100 and 1000-100 series, it seems that cross validation does better with more examples, but this falls directly from regularized linear regression doing better with more examples. Since you're using the same algorithm, if regularized linear regression does better with more examples, cross validation will do better with more examples.

Comparing the 100-10 and 100-100, it seems to be the case that when you have fewer features, cross validation doesn't give as much of an improved result. That

Figure 3: Regularized Linear Regression on 50(1000)-100

Figure 4: Regularized Linear Regression on 100(1000)-100

said, this is based on only one data point, and it could easily be luck that cross validation did not have a better result for the 100-10 data set. The evidence is not conclusive that number of features has an effect on the results of cross validation.

Refer to table 3 for the runtimes of the various problems for all of the data sets. It is immediately aparent that cross validation takes much more time than conventional linear regression. This is obvious from the algorithm - you're essentially doing the same calculation, but ten times instead of once. The results are superior, but at a steep computational cost.

4 Bayesian Linear Regression

If we examine the 1000-100 data set and its subsets, it is clear that Bayesian Linear Regression does better with more examples. The trend is clear and distinct.

It appears that Bayesian Linear Regression does better with fewer features, based on how much better the 100-10 data set's results were than the results from the 100-100 data set.

Table 1: Average runtime of 5 runs, in seconds			
Data Set	Regularized L.R.	Cross Validation	Bayesian L.R.
1000-100	5.813	31.730	0.135
50(1000)-100	2.297	4.720	0.071
100(1000)-100	2.325	6.247	0.0486
150(1000)-100	2.707	7.889	0.0540
100-10	2.229	1.937	0.0174
100-100	2.492	6.145	0.0417
wine	6.948	5.0199	0.0673

Figure 5: Regularized Linear Regression on 150(1000)-100

Figure 6: Regularized Linear Regression on wine

5 Comparison of Algorithms

As we've discussed earlier, Cross Validation is superior to Regularized Linear Regression for two reasons. Firstly, Cross Validation outperforms Regularized Linear Regression. Secondly, you only discover the optimal result of Regularized Linear Regression after you look at the answers from the test set.

Cross validation outperformed Bayesian Linear Regression in every case, as you can see in table 4. But, as we can see in table 3, Bayesian Linear Regression is orders of magnitude faster. Furthermore, if there are many examples, the results of Bayesian Linear Regression are comparable to Cross Validation. This implies that there are some cases where Bayesian Linear Regression may be the superior algorithm. If you require an algorithm that trains quickly on new data, and your data sources have many examples, then Bayesian Linear Regression is definitely the correct choice. But, if you have unlimited time to train your model, or your data set is small, it appears like Cross Validation would be the better algorithm.

Figure 7: Regularized Linear Regression on 100-100

Figure 8: Learning Curve for 1000-100 with Lambda = 20

Figure 9: Learning Curve for 1000-100 with Lambda = 80

Figure 10: Learning Curve for 1000-100 with Lambda = 140

Data Set	α	β	MSE	Cross Validation MSE	Target MSE
100-10	1.201	0.256	4.180	4.159	3.78
100-100	1.082	0.509	7.547	4.494	3.78
50(1000)-100	0.978	0.345	5.811	5.3	4.015
100(1000)-100	1.361	0.307	5.739	4.840	4.015
150(1000)-100	1.665	0.266	5.243	4.869	4.015
1000-100	2.733	0.264	4.338	4.137	4.015
wine	6.084	1.610	0.627	0.642	