COMP3314/CSIS0314: Assignment 1

Due on Monday, March 2, 2015

Instructor: Jack Wang

Qian Xin, 3035134147

2.1 Logistic regression

2.1.1 ML estimation

The effective learning rate found is 0.0002.
The effective maximum iteration values found is 2000.
Run the algorithm repeatedly for 3 times, the final accuracy is as follows,

	Training Set	Validation Set	Test Set
Big Set	1.00	0.82	0.90
	1.00	0.84	0.90
	1.00	0.85	0.86
Small Set	1.00	0.57	0.75
	1.00	0.61	0.61
	1.00	0.64	0.75

Table 1. Final accuracy of ML estimation, 3 times each set

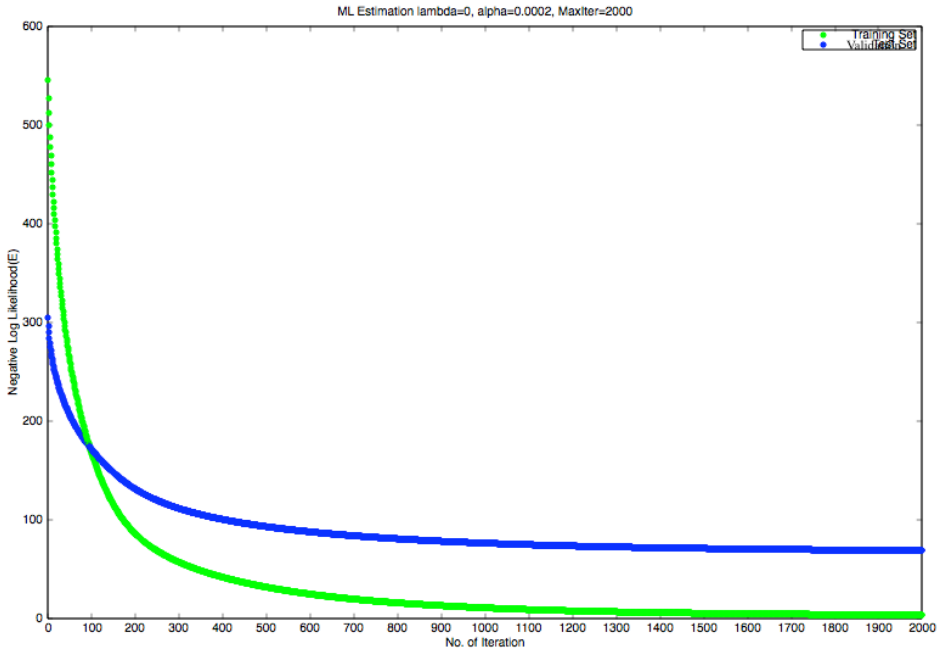

Figure 1. Plot on big set, accuracy = 1.00; 0.84; 0.90

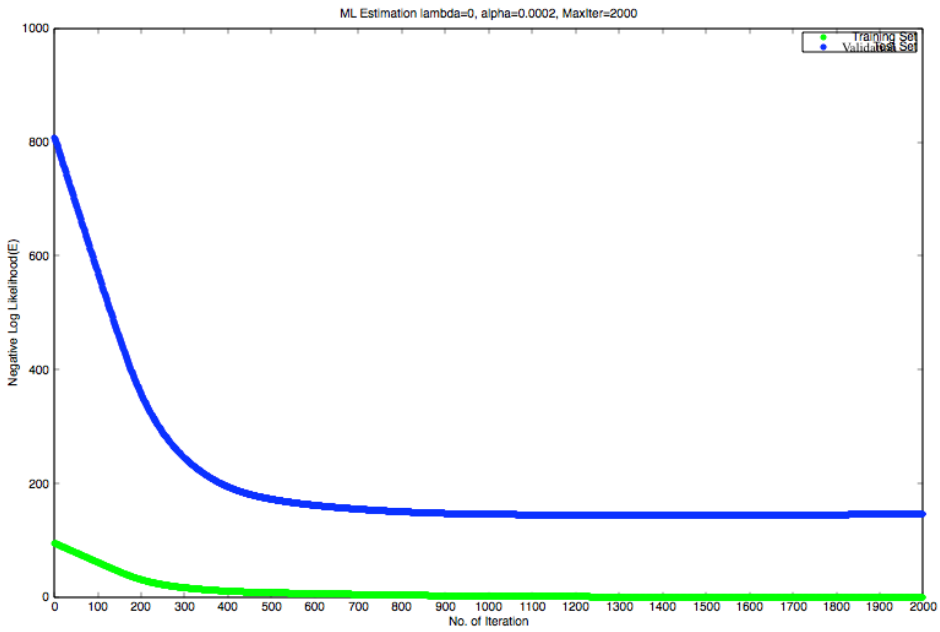

Figure 2. Plot on small set, accuracy = 1.00; 0.64; 0.75

Compare the plot from big set and small set, the small set has lower accuracy on validation set and test set. Given the 28*28 array of digit representation, the training data set is too small thus caused overfitting.

From Table 1, we may notice that the final accuracy and the plot have slight difference for each execution of the code. The source of the randomness is from the initialization of the weight vector w. The initial value of w is randomized from a normal distribution.

```
% Initialize the weight vector using samples from a normal distribution.
w = randn(M+1, 1);
```

2.1.2 MAP estimation

For negative log likelihood in the big data set, the curve for the training set goes up when Lambda increases. The curve for the validation set goes down when Lambda approaching 5, and goes up when Lambda=100.

For negative log likelihood in the small set, the curve for the training set stays low, (around 0), regardless of Lambda. The curve for the validation set goes down when Lambda approaching 5, and goes up when Lambda=100.

For accuracy in the big set, the curve for the training set has the value of 1.00 at Lambda=0,1,3,5, while goes down slightly down at Lambda=100. The curve for the validation set rises when Lambda increases ([0,1,3,5]) and reaches a peak when Lambda=5. When Lambda=100, the accuracy goes down slightly compared to when Lambda=5.

For accuracy in the small set, the curve for the training set keeps a value of 1.00. The curve goes down between Lambda=0-1 while rises until Lambda=5. The peak is still at Lambda=5. When Lambda=100, the accuracy goes down slightly compared to when Lambda=5.

From the four plots, we may notice the effective lambda value is 5.

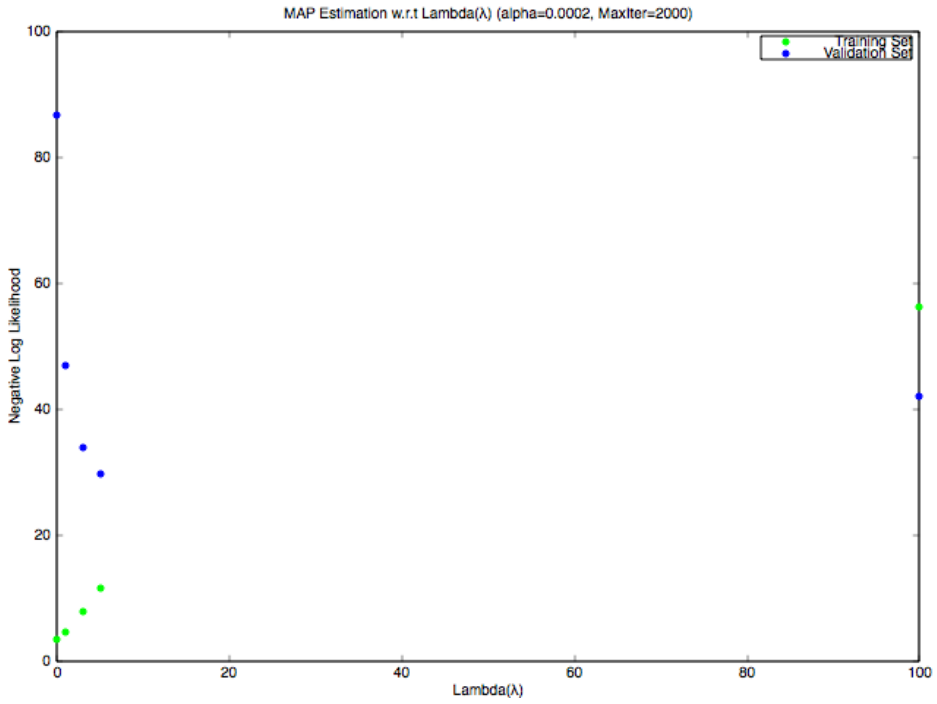

Figure 4. Average negative log likelihood of MAP estimation, on big set

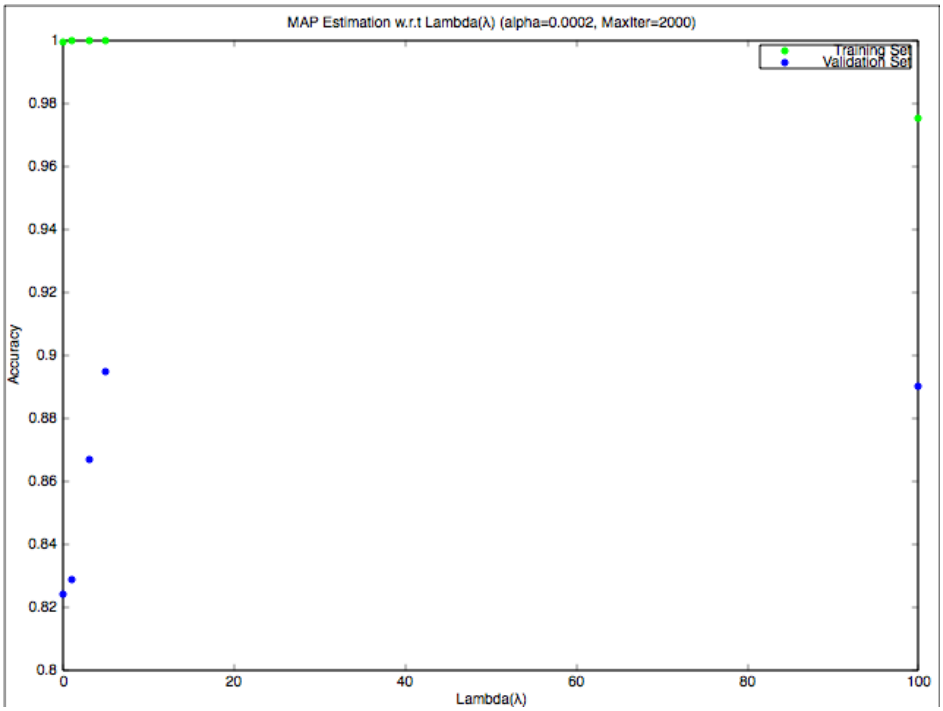

Figure 5. Average accuracy of MAP estimation, on big set

Given the classifier trained on the big set, the final accuracy on the test set is 0.93.

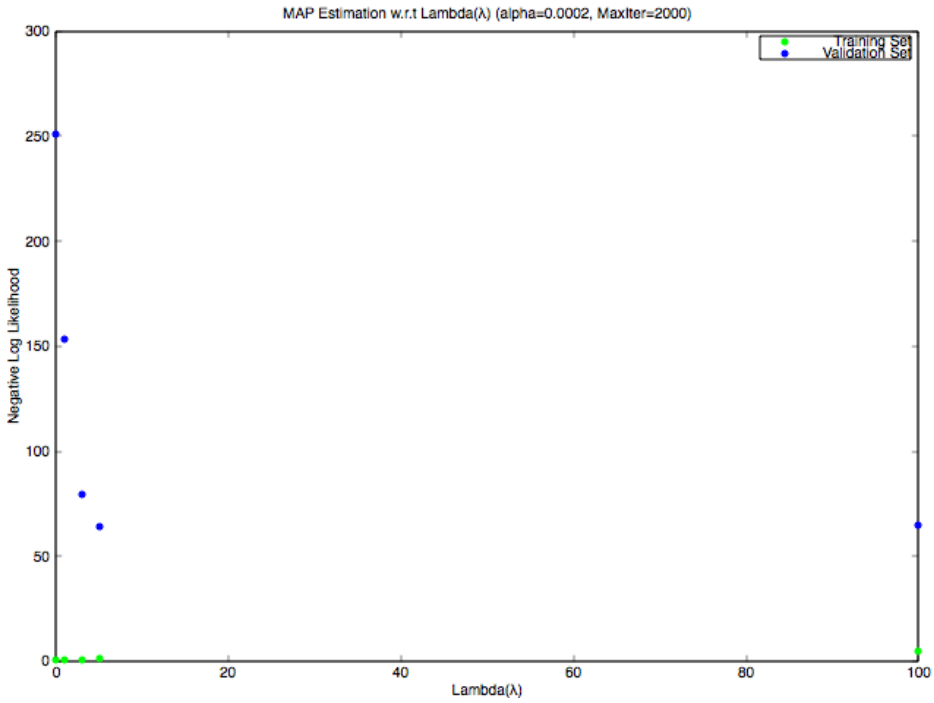

Figure 6. Average negative log likelihood of MAP estimation, on small set

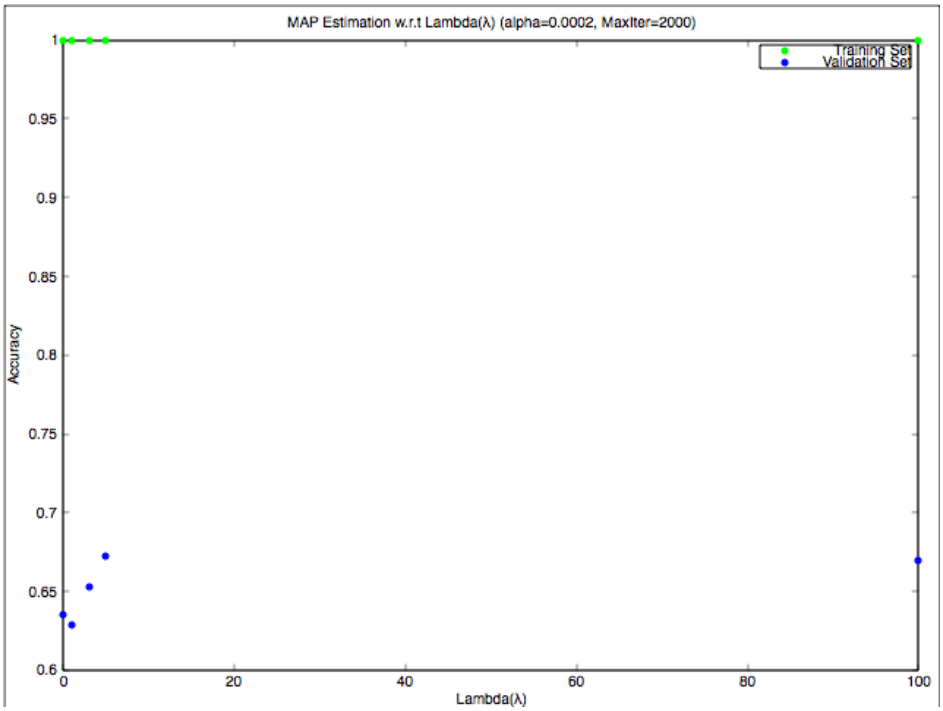

Figure 7. Average accuracy of MAP estimation, on small set

Given the classifier trained on the big set, the final accuracy on the test set is 0.81.

Because MAP introduces the hyperparameter Lambda and implements the regularization term, which controls the distribution of w. MAP estimation eliminates the problem of overfitting and results in a further rise on accuracy of the classifier.

On the full training set, the accuracy is raised from 0.90 to 0.93. On small training set, the raise on accuracy is more obvious, from 0.75 to 0.81. Such result is reasonable because the overfitting is a serious problem when the classifier is trained by smaller set. So that the regularization term has a more obvious effect on smaller training set.

2.2 k-NN classification

Figure 8 shows the accuracy of k-NN on the validation set and the test set.

From the validation set only, we may choose the most suitable value of k as 1, 9 or 13. Their accuracy on the test set is 0.94, 0.99 and 0.96. Combining the two results, we may choose k=9.

There is not much correlation between the accuracy of the validation set and the test set. We may say validation set is not able to predict test performance. This classifier is less accurate but more fast compared to logistic regression.

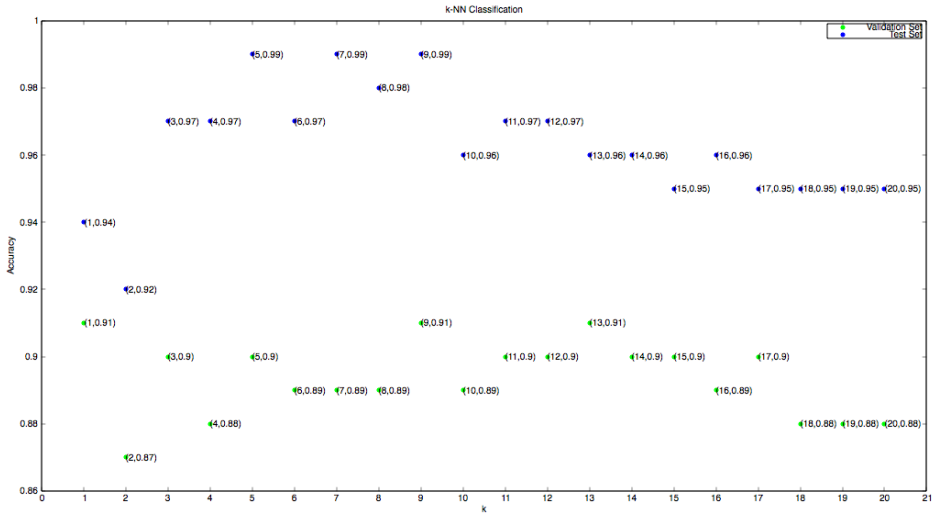

Figure 8. Accuracy of k-NN on both validation set and test set