Machine Learning CSE 574: Assignment 3

Group 6

Chintan Thakker - 5016 7864

Akash Mandole – 5016 8447

Pranshu Pancholi – 5016 9864

Results

Logistic Regression:

Logistic Regression is a probabilistic classifier. It is also a classic Discriminative Classifier model. In Discriminative classifier P(y|x) is modeled directly. Thus they can learn faster and with less data. Since it learns a Linear Boundary, Logistic Regression is said to be a linear classifier.

	Accuracy		
	Training Set	Validation Set	Test Set
Binary Logistic Regression	86.254%	85.47%	85.43%
Multi Class Logistic Regression	93.39%	92.43%	92.67%

Support Vector Machines:

Support vector machines is a hyperplane based classifier. It constructs a hyperplane such that the distance from it to the nearest data point on each side is maximized. In general the larger the margin the lower the generalization error of the classifier.

	Accuracy		
	Training Set	Validation Set	Test Set
Linear Kernel	97.286%	93.64%	93.78
RBF kernel and gamma = 1.0	100.0%	15.48%	17.14%
RBF kernel and default gamma	94.294%	94.02%	94.42%
RBF Kernel and C = 1	94.294%	94.02%	94.42%
RBF Kernel and C = 10	97.132%	96.18%	96.1%
RBF Kernel and C = 20	97.952%	96.9%	96.67%
RBF Kernel and C = 30	98.372%	97.1%	97.04%
RBF Kernel and C = 40	98.706%	97.23%	97.19%
RBF Kernel and C = 50	99.002%	97.31%	97.19%
RBF Kernel and C = 60	99.196%	97.38%	97.16%
RBF Kernel and C = 70	99.34%	97.36%	97.26%
RBF Kernel and C = 80	99.438%	97.39%	97.33%
RBF Kernel and C = 90	99.542%	97.36%	97.34%
RBF Kernel and C = 100	99.612%	97.41%	97.4%

From the above results we can see that SVM performs better than Logistic Regression, this is because logistic regression learns any hyperplane that correctly separates the data, while SVM learns the one which maximizes the distance from it to the nearest data point on each side (i.e. it learns the best one). A standard SVM is a type of linear classification using dot product. We can replace this dot product with kernel functions such as Gaussian radial basis function. Gamma is the free parameter of the Gaussian radial basis function. Gamma can be thought of as a regularization parameter. A high gamma means hyperplane is more prone to adjust too much to the training examples. Thus from the above results we can see that when gamma is too high, we are having very low accuracy on both validation and test set even if training set has 100% accuracy, this is because of overfitting. When gamma is 0 we have much better results on validation and test set.

Varying C Values:

To extend SVM to cases in which the data are not linearly separable, we introduce C or "Soft Margin" SVM parameter which allows some examples to be ignored or placed on the wrong side of the hyperplane. C introduces error penalty for stability. Now larger the value of C, more is the error penalty. This results in an improvement in the accuracy in the case of a training dataset but results in the hyperplane being over fitted to the training data set and we can see that the accuracy consistently drops in the test and validation datasets as we keep increasing the value of C.