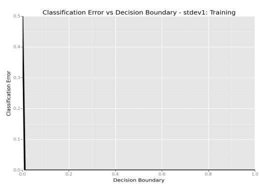
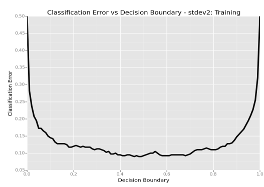
6.867 Problem Set 2

October 20, 2015

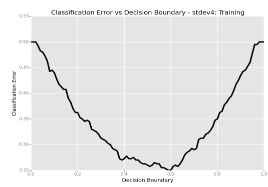
Logistic Regression



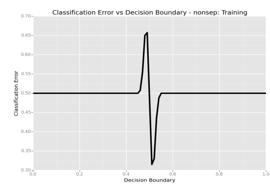
stdev1 (Training)



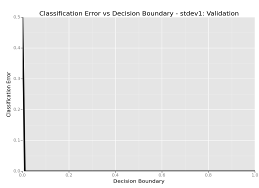
stdev2 (Training)



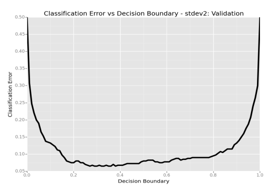
stdev4 (Training)



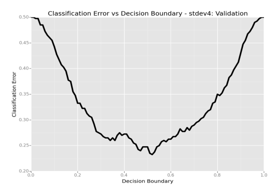
nonsep (Training)



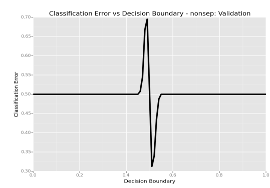
stdev1 (Validation)



stdev2 (Validation)



stdev4 (Validation)



nonsep (Validation)

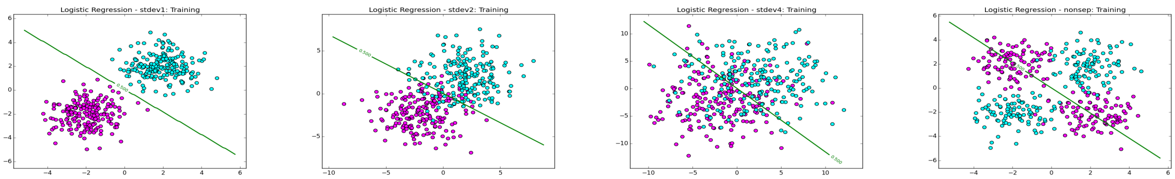
Figure 1: The classification error of Logistic Regression as a function of our choice of a decision boundary for the stdev1, stdev2, stdev4, and nonsep (training and validation) datasets

Support Vector Machines

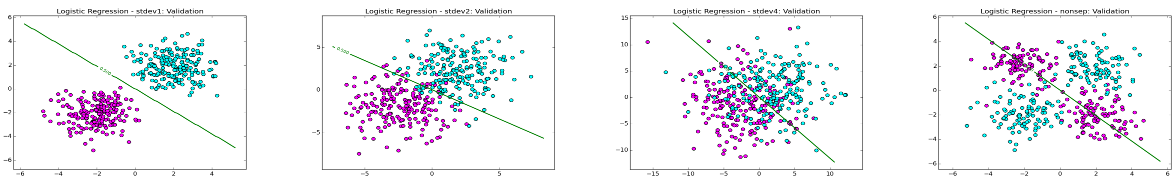
Let's now compare the performance of SVM to that of logistic regression for classification problems. To illustrate the objective and constraints of the support vector machine, we have included below the explicit objective we would optimize over, as well as the constraints, for the dual form of a linear SVM with slack variables. The equations below correspond to the 2D problem where we have positive examples (1, 2), (2, 2) and negative examples (0, 0), (-2, 3).

Table 1: Performance of Logistic Regression on provided data sets with a decision boundary of 0.5

dataset	classification error (training set)	classification error (validation set)
data_stdev1	0.00%	0.00%
data_stdev2	9.25%	8.00%
data_stdev4	26.00%	24.75%
data_nonsep	48.50%	50.75%



stdev1 (Training) stdev2 (Training) stdev4 (Training) nonsep (Training)



stdev1 (Validation) stdev2 (Validation) stdev4 (Validation) nonsep (Validation)

Figure 2: The decision boundaries generated by Logistic Regression plotted against the stdev1, stdev2, stdev4, and nonsep (training and validation) datasets

Table 2: Performance of Linear SVM on provided data sets

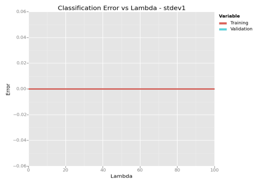
dataset	classification error rate (training set)	classification error rate (validation set)
data_stdev1	0.00%	0.00%
data_stdev2	9.50%	7.50%
data_stdev4	26.00%	23.50%
data_nonsep	49.50%	51.25%

$$\min_{\alpha_1,\alpha_2,\alpha_3,\alpha_4} \frac{1}{2} \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} 5 & 6 & 0 & -4 \\ 6 & 8 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ -4 & -2 & 0 & 13 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} + \begin{bmatrix} -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$$

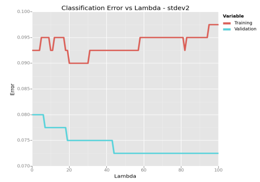
s.t.

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} \leqslant \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ C \\ C \\ C \\ C \end{bmatrix},$$

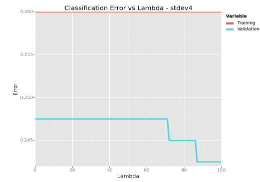
$$\begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = 0$$



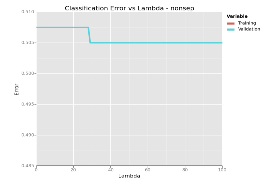
Classification Error - stdev1



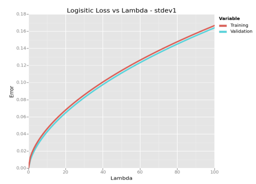
Classification Error - stdev2



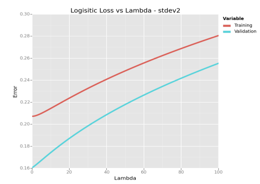
Classification Error - stdev4



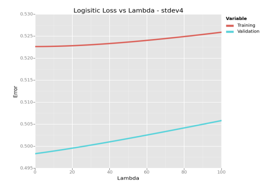
Classification Error - non-sep



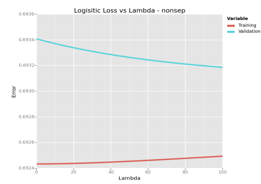
Logistic Loss - stdev1



Logistic Loss - stdev2

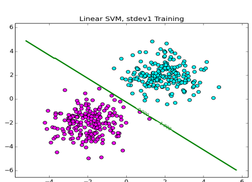


Logistic Loss - stdev4

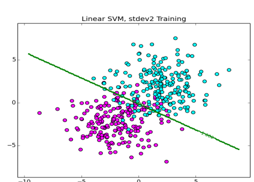


Logistic Loss - nonsep

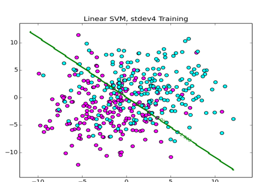
Figure 3: Classification Error (for a decision boundary of 0.5) and Logistic Loss of Logistic Regression as a function of the L2 penalty λ



stdev1 (Training)

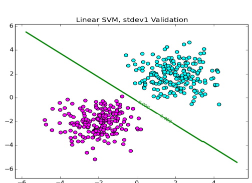


stdev2 (Training)

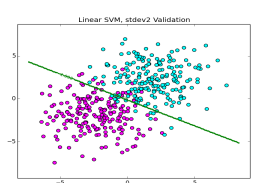


stdev4 (Training)

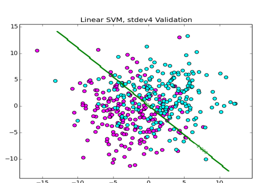
nonsep (Training)



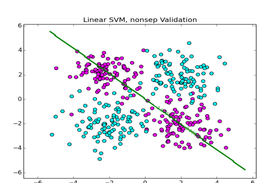
stdev1 (Validation)



stdev2 (Validation)



stdev4 (Validation)



nonsep (Validation)

Figure 4: The decision boundaries generated by SVM plotted against the stdev1, stdev2, stdev4, and nonsep (training and validation) datasets