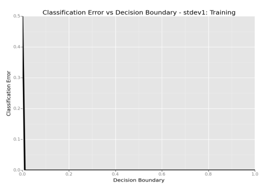
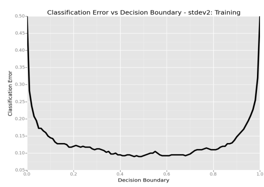
6.867 Problem Set 2

October 22, 2015

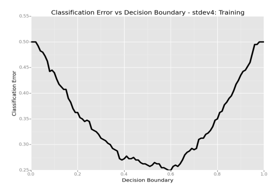
Logistic Regression



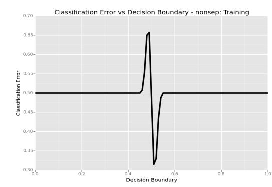
stdev1 (Training)



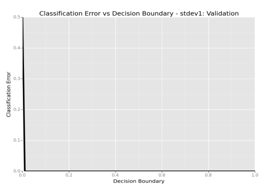
stdev2 (Training)



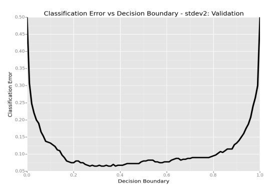
stdev4 (Training)



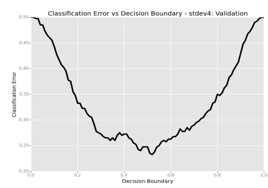
nonsep (Training)



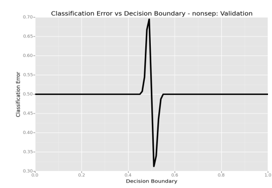
stdev1 (Validation)



stdev2 (Validation)



stdev4 (Validation)



nonsep (Validation)

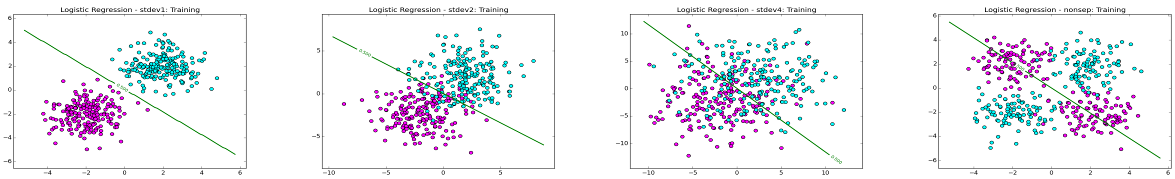
Figure 1: The classification error of Logistic Regression as a function of our choice of a decision boundary for the stdev1, stdev2, stdev4, and nonsep (training and validation) datasets

Support Vector Machines

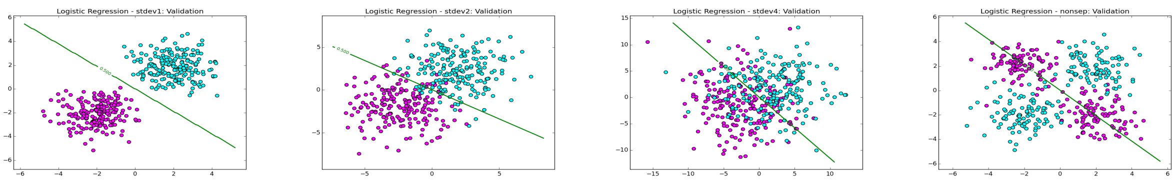
Let's now compare the performance of SVM to that of logistic regression for classification problems. To illustrate the objective and constraints of the support vector machine, we have included below the explicit objective we would optimize over, as well as the constraints, for the dual form of a linear 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%

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

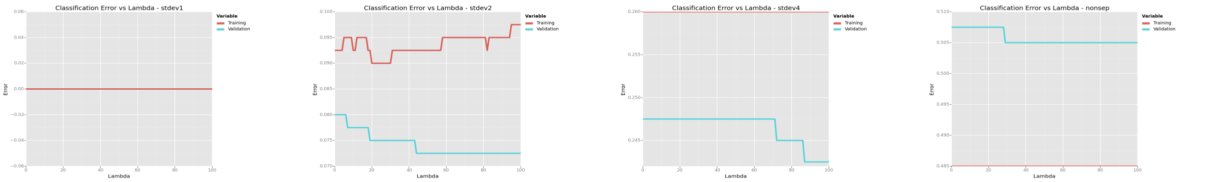
s.t.

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

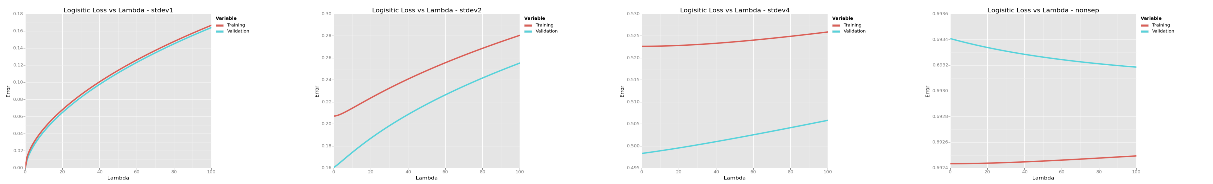
$$\begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = 0$$

Table 3: Soft SVM Performance for Different Parameters and Kernels

Kernel	C	σ	Number of Support Vectors	$1/\|\|\mathbf{w}\|\|$
Gaussian	.01	.01	400	.0286
Gaussian	.01	.1	400	.0286
Gaussian	.01	1	400	.0286
Gaussian	.01	10	400	.0286
Gaussian	.01	100	400	.0286
Gaussian	.1	.01	400	.0141
Gaussian	.1	.1	400	.0141
Gaussian	.1	1	400	.0207
Gaussian	.1	10	400	.0141
Gaussian	.1	100	400	.0141
Gaussian	1	.01	400	.0017
Gaussian	1	.1	391	.0022
Gaussian	1	1	100	.0076
Gaussian	1	10	368	.0018
Gaussian	1	100	400	.0017
Gaussian	10	.01	400	.0017
Gaussian	10	.1	390	.0018
Gaussian	10	1	71	.0017
Gaussian	10	10	209	.0003
Gaussian	10	100	400	.0002
Gaussian	100	.01	400	.0017
Gaussian	100	.1	390	.0018
Gaussian	100	1	59	.0002
Gaussian	100	10	159	7.51e-5
Gaussian	100	100	399	1.77e-5
Linear	.01	N/A	399	.0282
Linear	.1	N/A	393	.0133
Linear	1	N/A	392	.0017
Linear	10	N/A	393	.0002
Linear	100	N/A	397	1.8e-5

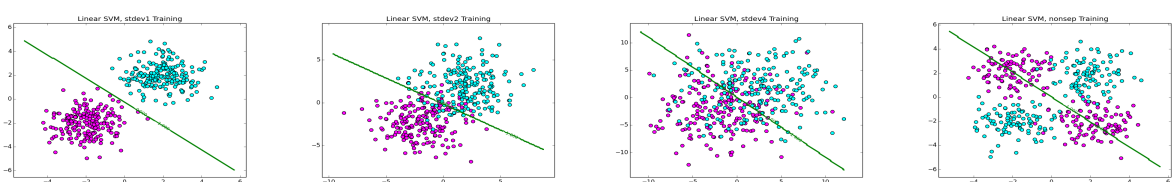


Classification Error - stdev1 | Classification Error - stdev2 | Classification Error - stdev4 | Classification Error - nonsep

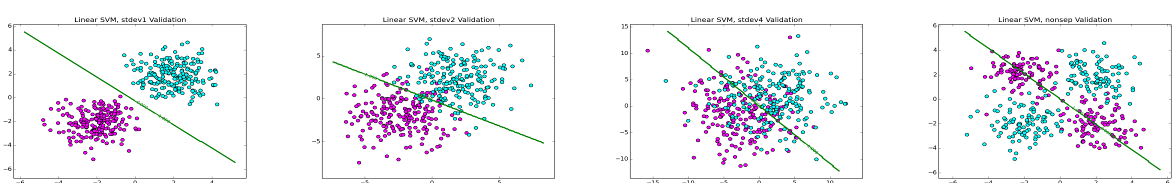


Logistic Loss - stdev1 | Logistic Loss - stdev2 | Logistic Loss - stdev4 | Logistic Loss - nonsep

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