6.867 Homework 2

Logistic Regression

1.1 Unregularized Logistic Regression

We started by implementing logistic regression with stochastic gradient descent and plotting the norm of the weight vector over time for both the unregularized (λ = 0) and L2 regularized (λ=1) objective.

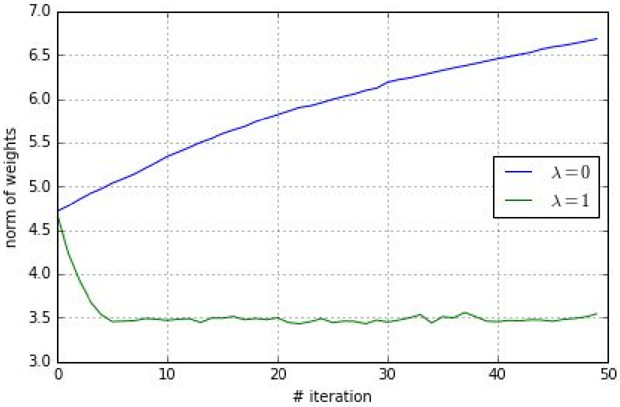

Figure 1. TODO: SOME DESCRIPTION.

As shown in Figure 1, the unregularized weight vector grows indefinitely while the regularized weight vector converges. This can be explained by the fact that larger magnitude weight vectors will result in more confident predictions, moving the prediction closer to 1.0 or 0.0 for positive and negative samples respectively.

Note, however, that a more confident predictor is not necessarily better. In fact, in many cases such as medical applications, a model that is both confident and incorrect is worse than a model that is uncertain on some inputs .

1.2 Hyperparameters for Logistic Regression

We examined each of the four datasets with various regularizers and λ values using the sklearn implementation of logistic regression. In our code, we explicitly adjust the intercept scaling so as not to regularize the bias term.

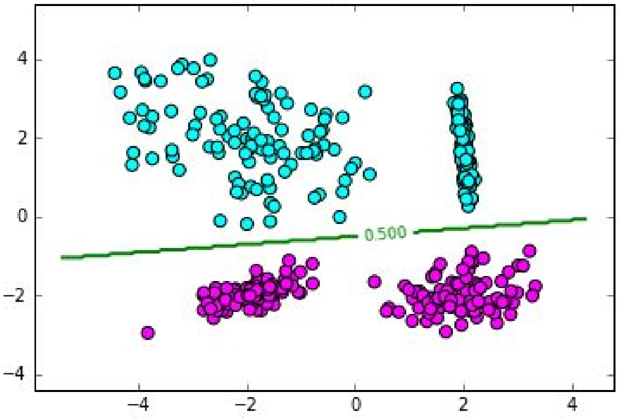

Figure 2. TODO: SOME DESCRIPTION.

The first dataset, shown in figure 2, is linearly separable and every combination of regularizer and λ between 0.0 and 20.0 produced a decision boundary which perfectly separated the two classes on the training, validation, and test set. As the λ constant increases, the norm of the weight vector shrinks.

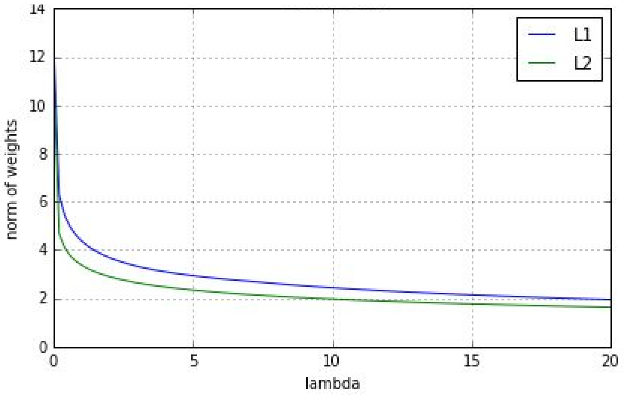

Figure 3. TODO: SOME DESCRIPTION .

On this particular dataset, we can completely ignore the first dimension of the input vector and still separate the two classes. This is reflected by the fact that with a large regularization constant, the L1 regularizer produces a sparse weight vector while the L2 regularizer produces a weight vector where the first dimension is near-zero.

The second dataset, however, is not linearly separable and there is no linear decision boundary capable of perfectly discriminating between positive and negative samples.

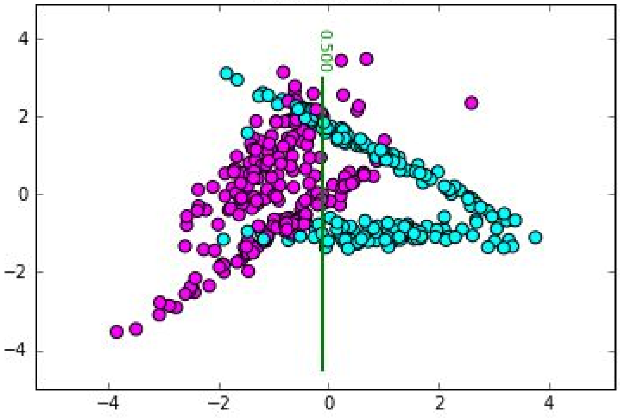

Figure 4. TODO: SOME DESCRIPTION.

As we increased λ, we observed that the norm of the weights behaved exactly like it did for the first dataset. On the other hand, the training and validation accuracy changed significantly as we adjusted λ and the regularizer.

A subset of the regularizer and λ values we examined is shown below in figure 5, where the bolded entry indicates the pair which minimizes the validation error.

L	λ	Train Error	Validation Error	Test Error
L1	0	0.1700	0.1750	0.1950
L2	0	0.1700	0.1750	0.1950
L1	**1**	**0.1675**	**0.1750**	**0.1950**
L2	1	0.1675	0.1750	0.1950
L1	2	0.1700	0.1800	0.1900
L2	2	0.1650	0.1750	0.1950

Figure 5. TODO: SOME DESCRIPTION

For the third dataset…

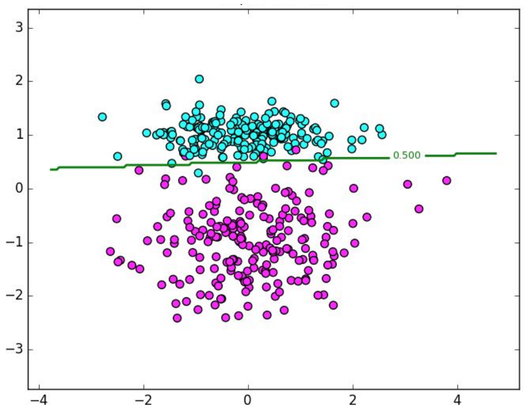

Figure 6. TODO: SOME DESCRIPTION.

L	λ	Train Error	Validation Error	Test Error
-	0	0.0125	0.035	0.050
L1	1	0.0175	0.035	0.045
L2	1	0.0225	0.030	0.030
L1	**2**	**0.0250**	**0.025**	**0.035**
L2	2	0.0250	0.035	0.030
L1	3	0.0275	0.035	0.030
L2	3	0.0275	0.040	0.030

Figure 7. TODO: SOME DESCRIPTION.

The fourth dataset is similar to the classic XOR problem and no combination of regularizer and λ worked...

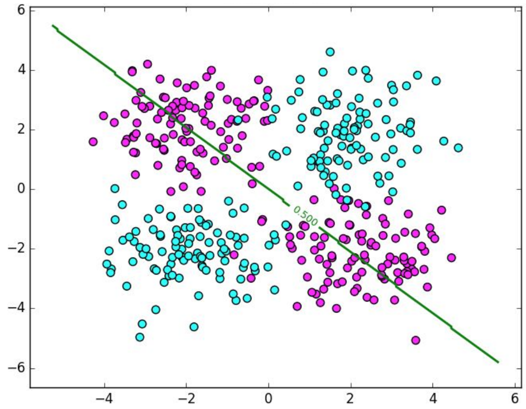

Figure 8. TODO: SOME DESCRIPTION.

L	λ	Train Error	Validation Error	Test Error
-	0	0.4850	0.5075	0.5025
L1	**1**	**0.4825**	**0.5025**	**0.5000**
L2	1	0.4850	0.5075	0.5025
L1	3	0.4825	0.5050	0.5000
L2	3	0.4850	0.5075	0.5025

Figure 9. TODO: SOME DESCRIPTION.

Support Vector Machine

Pegasos

MNIST