Support Vector Machines Analysis

Paul Laliberte' | CSCI-5622

1. Use the Sklearn implementation of svm's to train a classifier to distinguish 3's from 8's.

We implemented a Support Vector Machine with three different types of kernels: linear, polynomial, and radial basis function. The content to follow will discuss optimal parameters, classification performance on a hold-out set, and specific examples that were used as support vectors for each kernel.

2. Experiment with linear, polynomial, and RBF kernels. In each case, perform a grid search to determine optimal-ish hyperparameters for the given model.

For the grid search we used a logarithmic range from 10^{-9} to 10^{3} for all parameters. The number of samples used was 1000. The following table gives the optimal parameters.

	Linear	Poly	RBF
Optimal C	0.01	10.0	100.0
Optimal Degree	X	1	X
Optimal Gamma	X	X	.000316
Accuracy	.960	.961	.965

Figure 1: Optimal Parameters of Kernels

3. Comment on classification performance for each model for varying parameters by either testing on a hold-out set or performing cross-validation.

We have the following tables of classification reports and confusion matrices for each kernel. Both the classification report and confusion matrix are implementations from sklearn's metrics library. We used 1000 samples of the data, and tested each kernel three-to-four times with varying parameters.

First we give some definitions of the metrics. Precision is calculated as $\frac{\text{tp}}{(\text{tp}+\text{fp})}$, recall is calculated as $\frac{\text{tp}}{(\text{tp}+\text{fn})}$, f1-score is the weighted average of precision and recall, and support is the number of data points in the y_train set (labeled set). There are several things to notice. For the linear kernel we get a lower overall accuracy as we increased C. For the polynomial kernel increasing the degree by too great of a factor will produce a one-sided classification. We can counter this to an extent by increasing C, but for a degree 7 a C of 1000 is still too low. For the rbf kernel, we also see that there is an inverse relationship between the gamma and C. Increasing C requires a decrease in gamma, and vice versa.

Figure 2: Linear Kernel

C = 1.0	precision	recall	f1-score	support
3	0.94	0.96	0.95	1010
8	0.96	0.94	0.95	974
avg / total	0.95	0.95	0.95	1984

968	42
60	914

C = 100.0	precision	recall	f1-score	support
3	0.94	0.95	0.95	1010
8	0.95	0.93	0.94	974
avg / total	0.94	0.94	0.94	1984

964	46
65	909

C = 1000.0	precision	recall	f1-score	support
3	0.92	0.95	0.94	1010
8	0.94	0.92	0.93	974
avg / total	0.93	0.93	0.93	1984

956	54
78	896

Figure 3: Polynomial Kernel

C = 1.0, degree = 1.0	precision	recall	f1-score	support
3	0.96	0.95	0.95	1010
8	0.95	0.96	0.95	974
avg / total	0.95	0.95	0.95	1984

994	16
301	673

C = 1.0, degree = 3.0	precision	recall	f1-score	support
3	0.51	1.00	0.67	1010
8	0.00	0.00	0.00	974
avg / total	0.26	0.51	.34	1984

1010	0
974	0

C = 10.0, degree = 3.0	precision	recall	f1-score	support
3	0.82	.97	0.89	1010
8	0.96	0.77	0.86	974
avg / total	0.89	0.87	.87	1984

981	29
222	752

C = 1000, degree = 7.0	precision	recall	f1-score	support
3	0.51	1.00	0.67	1010
8	0.00	0.00	0.00	974
avg / total	0.26	0.51	.34	1984

1010	0
974	0

Figure 4: RBF Kernel

C = 1.0, gamma = 0.10	precision	recall	f1-score	support
3	0.98	0.97	0.97	1010
8	0.97	0.98	0.97	974
avg / total	0.97	0.97	0.97	1984

981	32
21	953

C = 1.0, gamma = 10	precision	recall	f1-score	support
3	0.51	1.00	0.67	1010
8	0.00	0.00	0.00	974
avg / total	0.26	0.51	.34	1984

1010	0
974	0

C = 100.0 , gamma = 10	precision	recall	f1-score	support
3	0.00	0.00	0.00	1010
8	0.49	1.00	0.66	974
avg / total	0.24	0.49	.32	1984

0	1010
0	974

C = 100.0, gamma = .0001	precision	recall	f1-score	support
3	0.96	0.97	0.97	1010
8	0.97	.96	0.96	974
avg / total	0.96	0.96	.96	1984

928	28
42	932

4. Give examples (in picture form) of support vectors from each class when using a linear kernel.

There were several support vectors for the linear kernel when using the entire dataset. The following figure shows three support vectors for the '3' class and three support vectors for the '8' class:

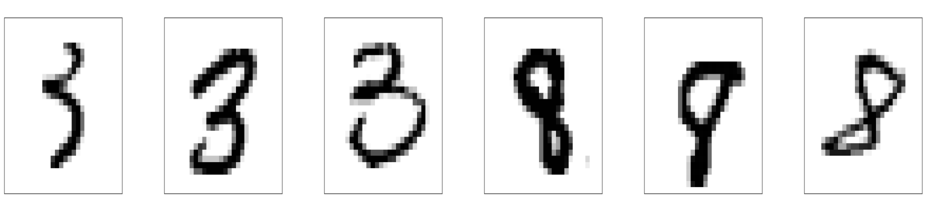

Figure 5: Support Vector Examples