

Foundations of Machine Learning — Homework Assignment 1

Anirudhan J Rajagopalan
N18824115
ajr619

November 08, 2015

C. Support Vector Machines

1

Installed the software from[1]. The installed version of software is also checked into github at[2].

2

See the following command:

```
$ ./svm-scale -s splice_noise_train.txt.range \  
> splice_noise_train.txt > splice_noise_train.txt.scale  
$ ./svm-scale -r splice_noise_train.txt.range \  
> splice_noise_test.txt > splice_noise_test.txt.scale
```

3

Run training and test script[3] by editing the KERNEL_DEGREE parameter for each value of $d = 1, 3, 5$.

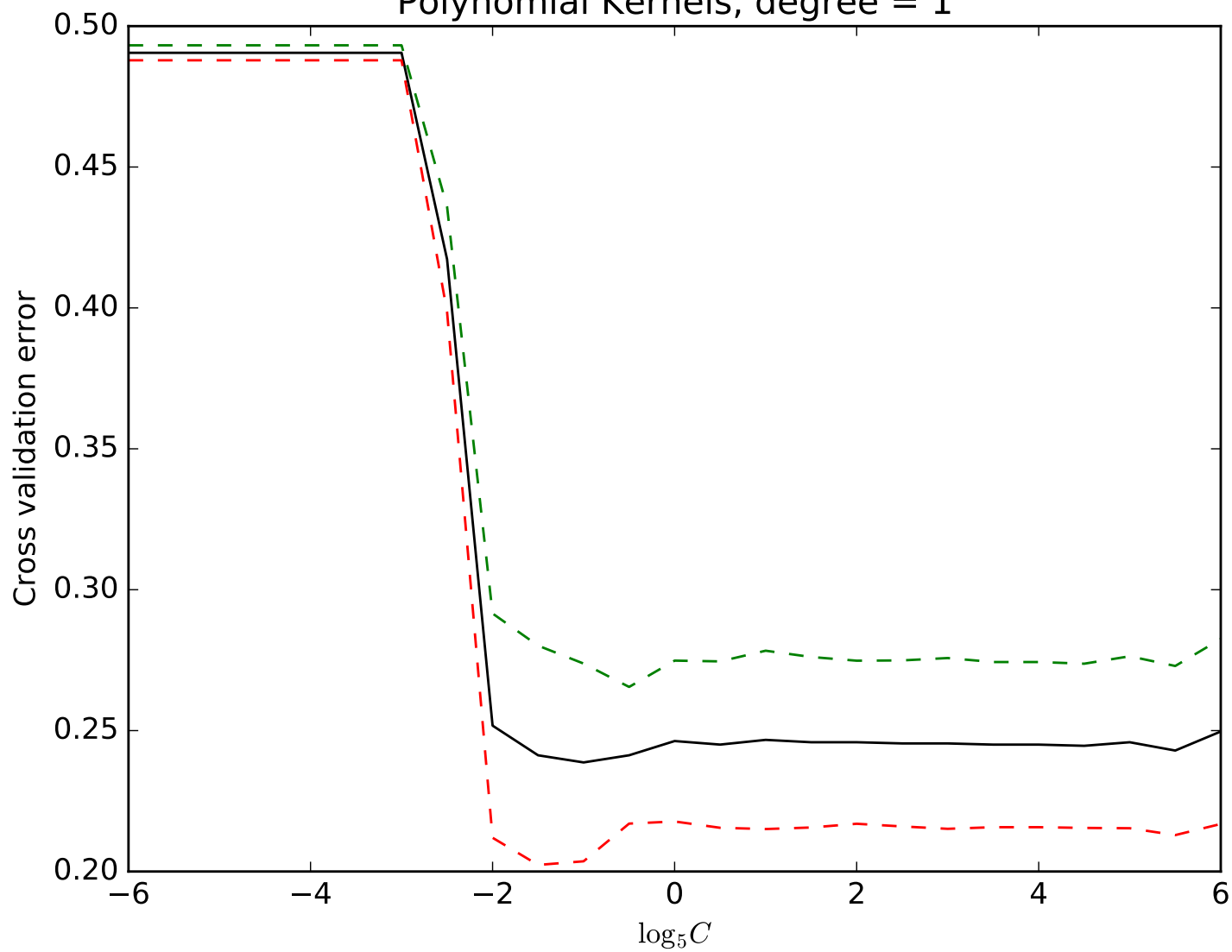
```
$ python cross_validation.py > deg1.out # KERNEL_DEGREE = 1  
$ python cross_validation.py > deg3.out # KERNEL_DEGREE = 3  
$ python cross_validation.py > deg5.out # KERNEL_DEGREE = 5
```

Filter the parameter and accuracy information from the run logs (deg1.out, deg3.out and deg5.out) by the command below.

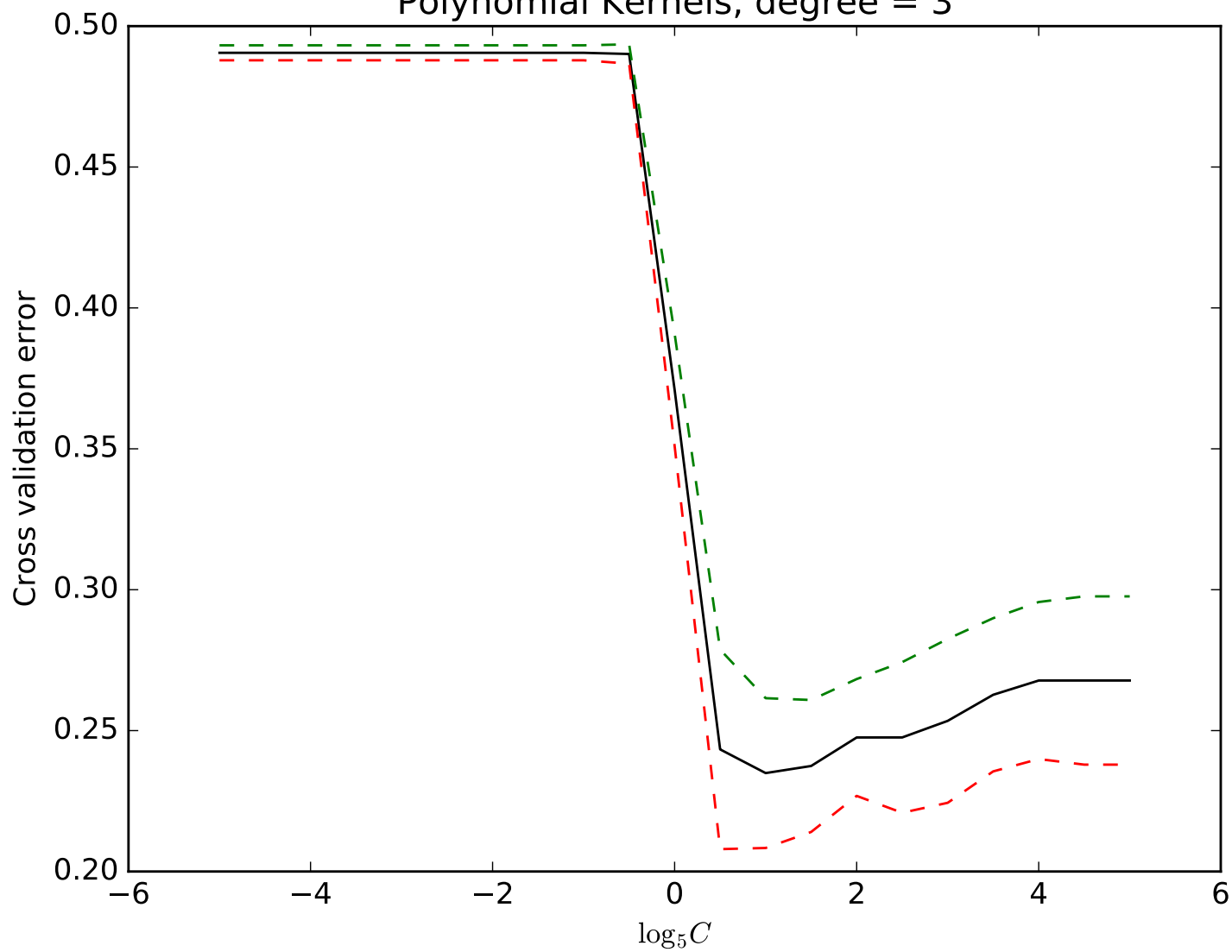
```
$ cat deg1.out | grep OUR | cut -d' ' \  
> -f2,3,4,5,6 > deg1.out.filtered  
$ cat deg3.out | grep OUR | cut -d' ' \  
> -f2,3,4,5,6 > deg3.out.filtered  
$ cat deg5.out | grep OUR | cut -d' ' \  
> -f2,3,4,5,6 > deg5.out.filtered
```

Use plotter.py[4] to create plots from the output values for KERNEL_DEGREE values 1, 3, 5. The output will be saved as deg1.pdf, deg3.pdf and deg5.pdf. All the three plots are embedded below one after the other. The dotted red line corresponds to -1 standard deviation. The dotted green line corresponds to +1 standard deviation and the black continuous line corresponds to the Mean.

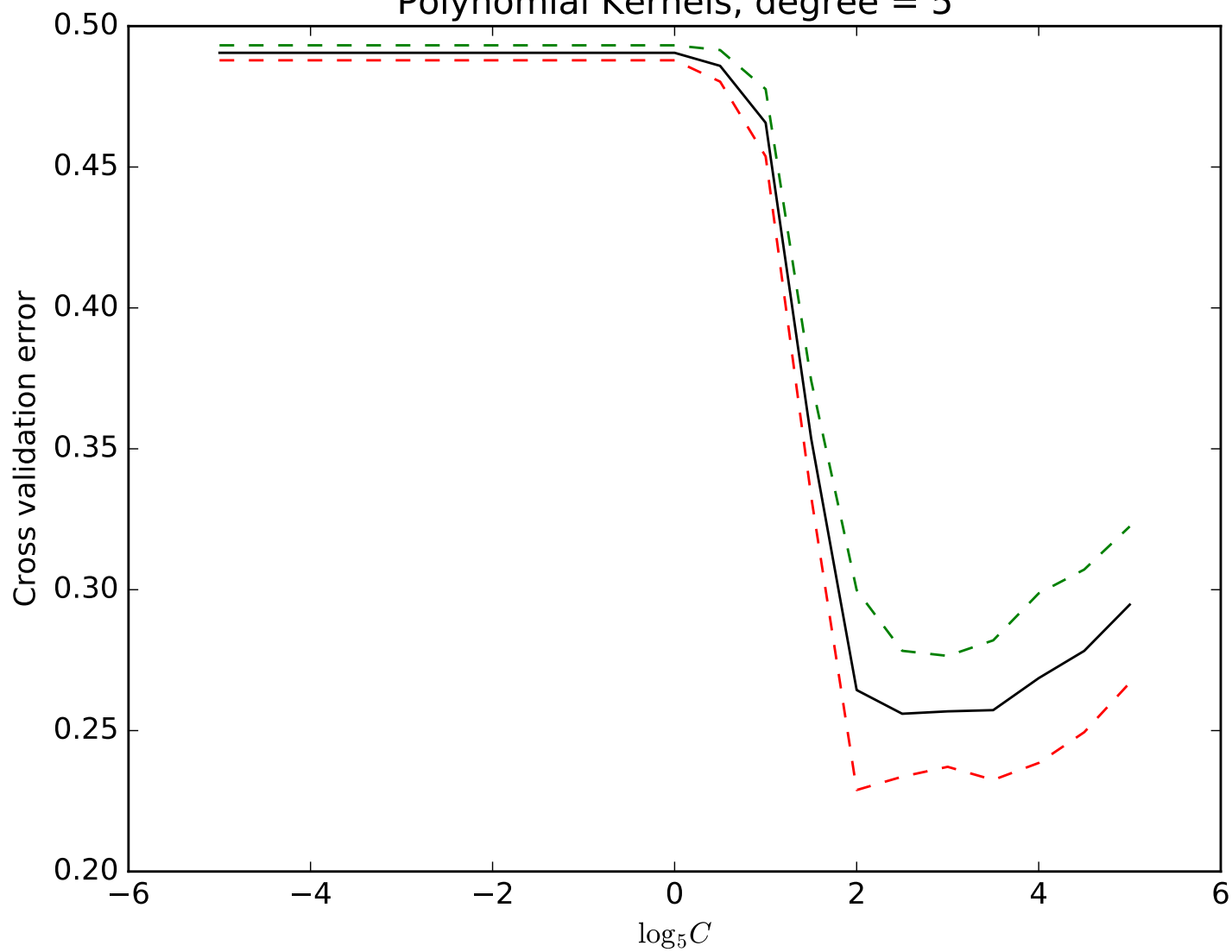
Polynomial Kernels, degree = 1



Polynomial Kernels, degree = 3



Polynomial Kernels, degree = 5



Best values of c for polynomial kernels 1, 3, 5 are:

$d = 1$	$c^* = 5^{-1.0} = 0.2$	$cv - err = 23.87\% \pm 3.5\%$
$d = 3$	$c = 5^{1.0} = 5$	$cv - err = 23.49\% \pm 2.65\%$
$d = 5$	$c = 5^{2.5} = 55.9017$	$cv - err = 25.59\% \pm 2.23\%$

The best C measured in the cross-validation set is $C^* = 5^{1.0}$ with degree $d^* = 3$ which gives an average error of $23.49\% \pm 2.65\%$

4

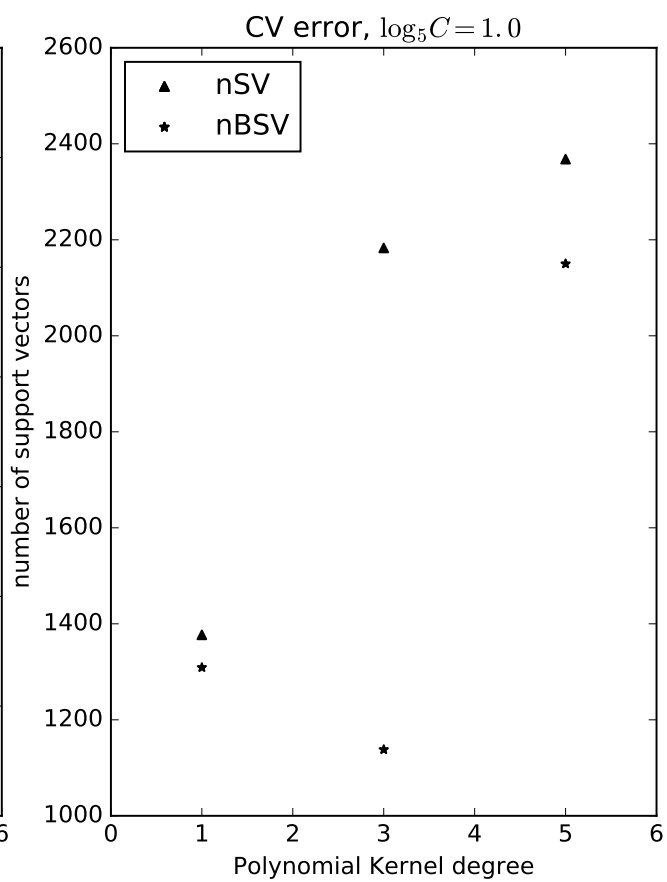
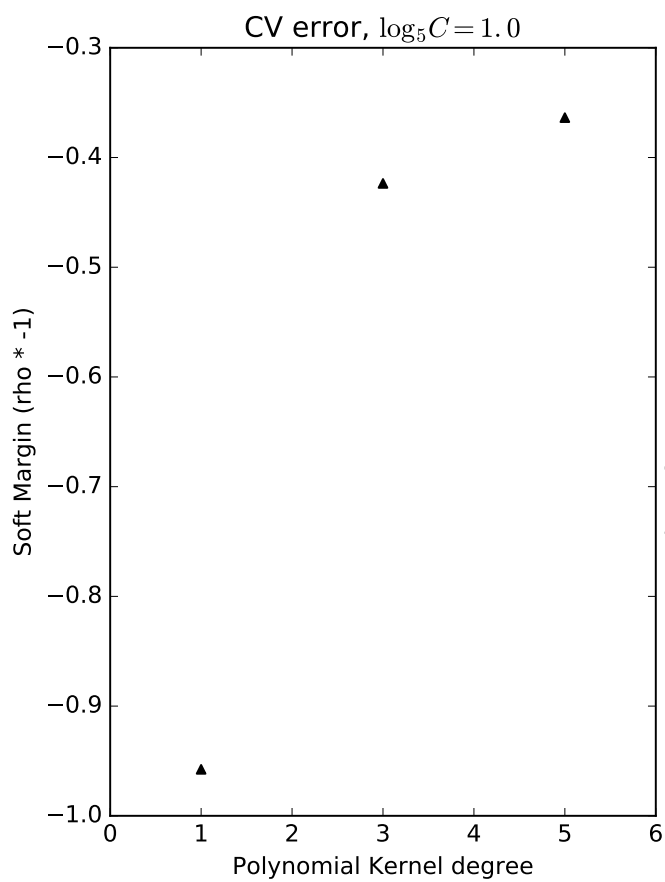
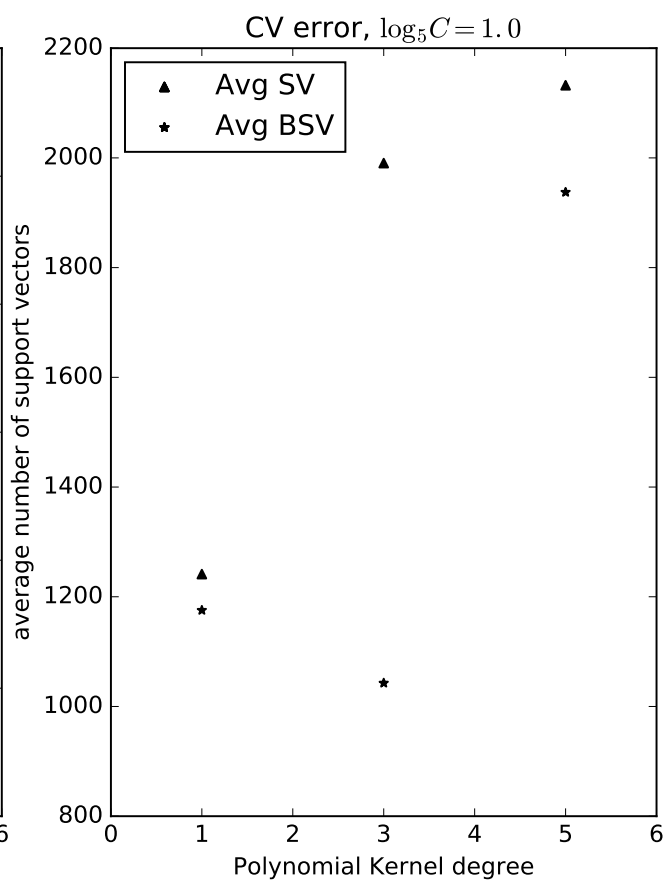
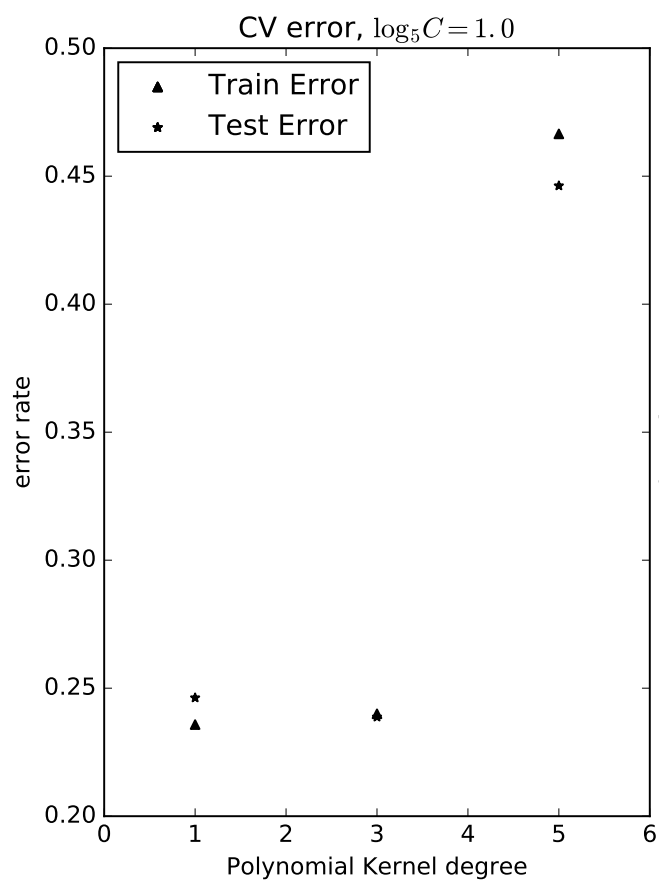
The average number of SV and BSV are taken across the cross validation sets. NumMarginal represents the average number of marginal support vectors.

Similarly nSV and nBSV represent the number of support vectors and bounded support vectors in test set. Marginal represents the number of support vectors in test set.

After manually populating d , cross-train accuracy, test accuracy, nSV, nBSV, rho, totalSV, totalBSV into a file[5], we run a script[6] to generate the plots as a pdf file.

D	Avg Sv	Avg BSV	numMarginal	nSv	nBSV	marginal
1	1241.3	1175.4	65.9	1377	1309	68
3	1990.5	1042.7	947.8	2183	1138	1045
5	2132.4	1937.4	195.0	2368	2150	218

Soft margin is given by the rho value. Svm-train outputs negative margin values (rho). We use rho values to plot the soft margins in our plots. The plots generated are as shown below.



5

We can modify grid.py to use 'C' in powers of 5 (change $2^{**}c$ to $5^{**}c$ in lines 142, 262, 348, 453 and 454). Running grid.py with $-\log 5c -4,4,1$ and $-\log 2g -4,4,1$ -v 10 -t 2, -s 0 gives us the best value of c and g.

We get the best cross validation accuracy of 79.7474 for $c = 1.0$ and $g = 0.03125$. Note that $\sigma = \sqrt{\frac{1}{2g}}$

We train the svm with the obtained c and v values

```
$ ./svm-train -t 2 -s 0 -c 1.0 -g 0.03125 \
> ./splice_hw/splice_noise_train.txt.scale ./rbf.model
$ ./svm-predict \./splice_hw/splice_noise_test.txt.scale \
> rbf.model rbf.output
```

gives us Accuracy = 80.125% (641/800) (classification)

The best accuracy we got with polynomial kernel SVM is 76.125 for $C^* = 5^{1.0}$ with degree $d^* = 3$.

The value of soft margin (rho) for the rbf kernel SVM is -0.586184.

6

D. Kernels

1

Given: Kernel, K is defined by $K(x, y) = \sum_{i=1}^N \cos^n(x_i^2 - y_i^2)$ for all $(X, Y) \in \mathbb{R}^N \times \mathbb{R}^N$

Solution: We know that

$$\cos(x_i^2 - y_i^2) = \sin(x_i^2) \cdot \sin(y_i^2) + \cos(x_i^2) \cdot \cos(y_i^2) \quad (1)$$

This can be written as a dot product of two vectors

$$\phi(x_i) = \begin{bmatrix} \cos(x_i^2) \\ \sin(x_i^2) \end{bmatrix} \quad \text{and} \quad \phi(y_i) = \begin{bmatrix} \cos(y_i^2) \\ \sin(y_i^2) \end{bmatrix} \quad (2)$$

K can be written as summation of N kernels that are a product of n K' kernels. $K' = \cos(x_i^2 - y_i^2)$. Hence to prove that K is PDS, it is enough to show that K' is PDS.

To show K' is PDS, we show that $c^T K' c > 0$ for any column vector $C = (C_1, C_2, \dots, C_m)^T \in \mathbb{R}^m$

$$C^T K' C = \sum_{i,j=1}^m c_i c_j \begin{bmatrix} \cos(x_i^2) \\ \sin(x_i^2) \end{bmatrix} \begin{bmatrix} \cos(y_j^2) \\ \sin(y_j^2) \end{bmatrix} \quad (3)$$

$$= \sum_{i=1}^n c_i \begin{bmatrix} \cos(x_i^2) \\ \sin(x_i^2) \end{bmatrix} \times \sum_{j=1}^n c_j \begin{bmatrix} \cos(y_j^2) \\ \sin(y_j^2) \end{bmatrix} \quad (4)$$

$$= \left[\sum_{i=1}^n c_i \begin{bmatrix} \cos(x_i^2) \\ \sin(x_i^2) \end{bmatrix} \right]^2 \quad (5)$$

$$\geq 0 \quad (6)$$

Which proves K' is PDS.

Hence $K(x, y) = \sum_{i=1}^N \cos^n(x_i^2 - y_i^2)$ is PDS.

1

Given: $K(x, y) = e^{-\frac{\|x-y\|}{\sigma}}$

and $\|x - y\| = \frac{1}{2\Gamma\frac{1}{2}} \int_0^\infty \frac{1 - e^{-t(\|x-y\|^2)}}{t^{3/2}} dt$ valid for all x, y

To prove: K is PDS

Proof: A normalized kernel K' for any kernel K is given by

$$K' = \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}}$$

For a PDS kernel given by $K_1 = e^{2txy} \forall t \geq 0$ we have the normalized kernel K_1 as

$$= \frac{e^{2txy}}{\sqrt{e^{2tx.x} e^{2ty.y}}} \quad (7)$$

$$= \frac{e^{2txy}}{e^{t(x.x+y.y)}} \quad (8)$$

$$= e^{-t(\|x-y\|^2)} \quad (9)$$

Since K_1 is PDS, its normalized kernel K_1 is also a PDS kernel.

$\Rightarrow -e^{-t(\|x-y\|^2)}$ is NDS and also $1 - e^{-t(\|x-y\|^2)}$ is also NDS. Since $t \geq 0$, we can say that $K_{nds} = \frac{1 - e^{-t(\|x-y\|^2)}}{t^{3/2}}$ is also a NDS kernel.

Now, for the kernel $K_2 = \frac{1}{2\Gamma\frac{1}{2}} \int_0^\infty \frac{1 - e^{-t(\|x-y\|^2)}}{t^{3/2}} dt$ for any column vector

$$C = (C_1, C_2, \dots, C_m)^T \in \mathbb{R}^m$$

$$\begin{aligned} C^T K_2 C &= \sum_{i,j=1}^n c_i c_j K_2(x_i x_j) \\ &= \sum_{i,j=1}^n c_i c_j \left[\frac{1}{2\Gamma\frac{1}{2}} \int_0^\infty \frac{1 - e^{-t(\|x-y\|^2)}}{t^{3/2}} dt \right] \\ &= \frac{1}{2\Gamma\frac{1}{2}} \left[\int_0^\infty \sum_{i,j=1}^n c_i c_j \frac{1 - e^{-t(\|x-y\|^2)}}{t^{3/2}} dt \right] \end{aligned}$$

Since we know that $K_{nds} \leq 0$;

$$C^T K_2 C \leq 0$$

$\Rightarrow K_2$ is also NDS.

This proves that $\|x - y\|$ is an NDS kernel. Since $\|x - y\|$ is NDS, from lecture5 which states that K is NDS iff $\exp(-tK)$ is PDS for all $t > 0$:

$\Rightarrow e^{-t\|x-y\|}$ is a PDS kernel where $t = \frac{1}{\sigma}$

References

- [1] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [2] <http://git.io/v80yn>
- [3] <http://git.io/v80yY>
- [4] <http://git.io/v80yk>
- [5] <http://git.io/v8Edr>
- [6] <http://git.io/v8EdH>
- [7] <http://git.io/v8KT1>