T-61.6040 Assignment-04/2011

Santosh Tirunagari (245577)
santosh.tirunagari@aalto.fi

October 17, 2011

AIM

Implemention of support vector data description algorithm discussed in the lecture using Gaussian and linear kernel for a toy dataset.

$$k_{GAU} = (x_i, x_j) = exp(-\|x_i x_j\|_2^2/s^2),\ and\ s = \{16, 32, 64, 128\}. \quad (1)$$

$$k_{Lin} = (x_i, x_j) =< x_i, x_j > . \quad (2)$$

Datasets

The datasets for hand written digits used in the assignment are training set with 256 features and a classlabel set. Class labels set are only used to mark the scatter plot.

Implementation

Load and zscore normalize the X. Zscore normalization can also be done using the method described in the lecture.

```
 X = zscore(dlmread('usps_012_X.txt'));
Y = (dlmread('usps_012_y.txt'));
```

Calculate Gaussian kernel function.

```
K = exp(-pdist2(X,X).^2/S(i).^2);
```

Center the kernal and solve eigenvalue problem.

```
    Kc1 = Kc1./m;
Kc = K-(Kc1*K)-(K*Kc1)+(Kc1*K*Kc1);
[U, lambda]= eigs(Kc, 2, 'lm');
```

Normalize the eigenvectors.

```
lambda = sqrt(diag(lambda));
U1 = U(:,1)./lambda(1);
U2 = U(:,2)./lambda(2);
U = [U(:,1) U(:,2)];
```

Project the mean centered kerner matrix on to the eigenvectors.

```
pom = Kc*U;
```

Plot the scatterplot

```
figure(i);
scatter(pom(:,1),pom(:,2),1,Y);
for k=1:length(pom)
    if(Y(k) ==0)
        text(pom(k,1),pom(k,2),num2str(Y(k)),'color','r');
    elseif(Y(k) ==1)
        text(pom(k,1),pom(k,2),num2str(Y(k)),'color','g');
    else
        text(pom(k,1),pom(k,2),num2str(Y(k)),'color','b');
    end
end
axis tight;
```

save the plots

Plot the graphs showing different hyperspheres.

```
title(['K_Gau S = ',num2str(S(i))]);
  %---print figure---
  fnout = ['S=', num2str(i,'%.2d'), '.jpg'];
  print('-djpeg','-r150',fnout);
```

The Gaussian kernel implementation is carried out for all the values of $s = \{16, 32, 64, 1288\}$ and these plots are shown in figure 1.

similarly the same implementation is carried out for linear kernel the plot is shown in figure 2.

Running of code for Gaussian kernel and linear kernel in matlab are 'KPCA'.

Results

Figure 1 shows that the S value increases the guassian kernel behaves more like a linear kernal. With the smaller S value the small Gaussian distribution is actually seen properly. As the S increases the width of the Gaussian distribution tends to increase.

Figure 1: Scatter plot showing Gaussian kernel.

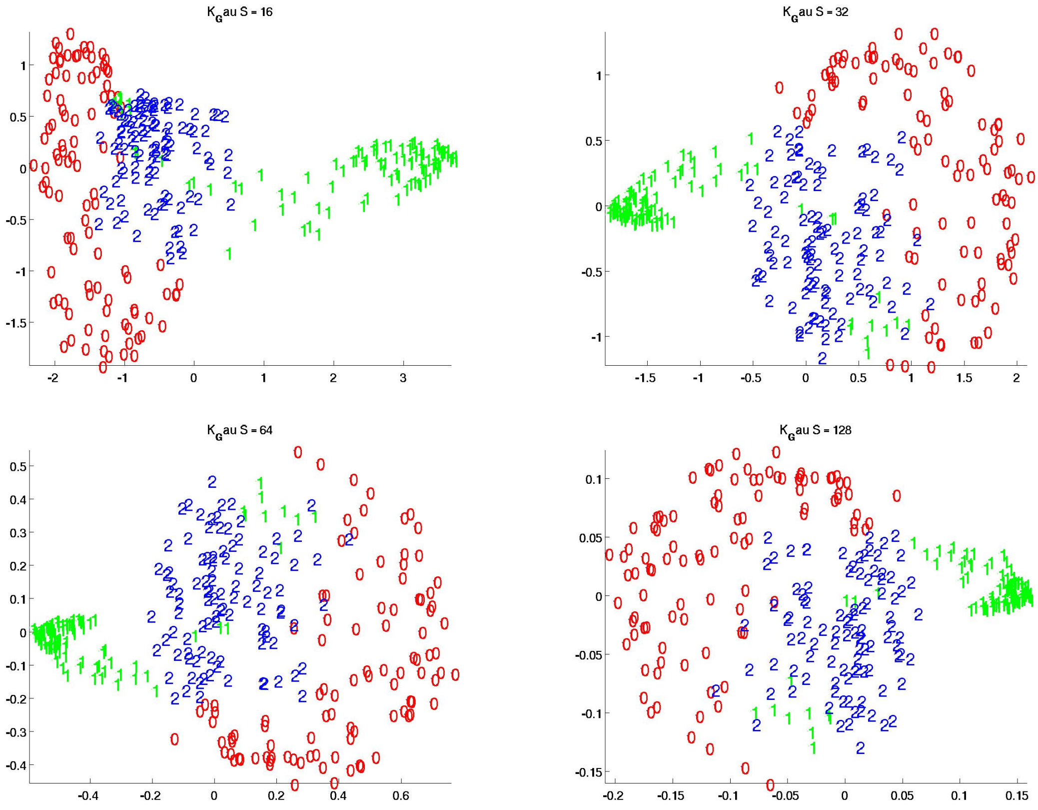
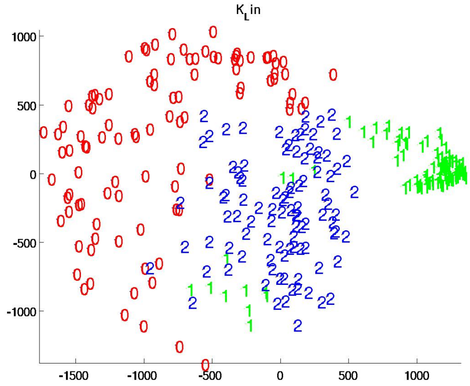


Figure 2 shows the linear kernel whole data in 2 dimensions. Figure 1, the σ value is significantly large the guassian kernel behaves more like a linear kernal.

Figure 2: Scatter plot showing linear kernel.



References

[1] Lecture slides and course book.

Attachments

Matlab code KPCA.m and datasets have been archived along with this report.