EC 414 Midterm 2 Practice Questions

Prof. Brian Kulis

April 4, 2019

1 SVM

1) Suppose we have 2 points that belong to Class 1: $\boldsymbol{x}_1 = [1,2]^T$, $\boldsymbol{x}_2 = [1,4]^T$, and 1 point that belongs to Class 2: $\boldsymbol{x}_3 = [3,4]^T$. Suppose we are given the parameters of a hyperplane: $\boldsymbol{w} = [2,1]^T$, $b = -6$. Is this an SVM hyperplane? Explain.

2) Describe some pros and cons of SVM.

2 Clustering

1) Can the polynomial kernel $K(\boldsymbol{x}_1, \boldsymbol{x}_2) = (\boldsymbol{x}_1^T \boldsymbol{x}_2 + b)^2$ correctly cluster the two circle data shown below? Explain.
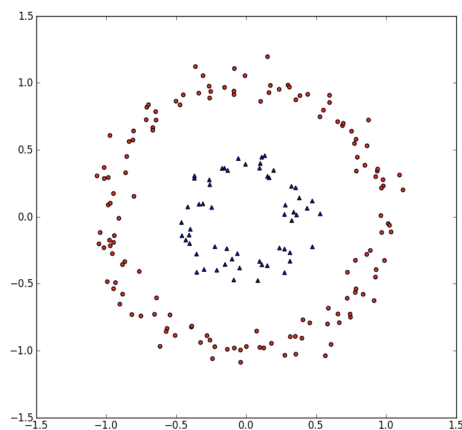

Figure 1: Circle data.

2) Imagine we use an alternative distance metric for the assignment step in k-means that factors in knowledge of the variance structure of a given cluster. Specifically, we use the Mahalanobis distance between a point $\boldsymbol{x}_i$ and a cluster with parameters $\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j$ as the metric: $d_{\mathrm{mahalanobis}} = \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_j)}$.

Write out the full adjusted k-means algorithm to account for this alternate distance metric.

3   PCA

1) Calculate the eigenvectors and eigenvalues of the following matrix:

$$M = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

2) In PCA, does mean-centering the data give a different result compared to that of not mean-centering the data at all? Explain.