Homework 3

Name: Linpeng Lyu
Andrew ID: linpengl

2 Kernel PCA

1. Solution:
   Because any principal component $\boldsymbol{w}$ is the eigenvector of $\boldsymbol{X} \cdot \boldsymbol{X^T}$, where columns of $\boldsymbol{X}$ are the n data points. So,

$$(\boldsymbol{X} \cdot \boldsymbol{X}^T) \cdot \boldsymbol{w} = \lambda \boldsymbol{w} \tag{1}$$

   where λ is the eigenvalue of matrix $\boldsymbol{X} \cdot \boldsymbol{X^T}$ corresponding to eigenvector $\boldsymbol{w}$. So we can get that:

$$\begin{aligned} \boldsymbol{w} &= \frac{\boldsymbol{X} \cdot \boldsymbol{X}^T \cdot \boldsymbol{w}}{\lambda} \\ &= \boldsymbol{X} \cdot \frac{\boldsymbol{X}^T \cdot \boldsymbol{w}}{\lambda} \end{aligned} \tag{2}$$

   Set $\boldsymbol{a}$ to $\frac{\boldsymbol{X}^T \cdot \boldsymbol{w}}{\lambda}$, so we can get the following equation from equation (2).

$$\boldsymbol{w} = \boldsymbol{X} \cdot \boldsymbol{a} \tag{3}$$

   Thus, any principal component $\boldsymbol{w}$ can be written as a linear combination of the data points $\{\boldsymbol{x}_i\}_{i=1}^n$.

2. Solution:
   From the lecture, I can get the objective of maximizing variance between projected data points (1^st PC vector).

$$\max_{\boldsymbol{w}} \boldsymbol{w}^T (\boldsymbol{X} \cdot \boldsymbol{X}^T) \boldsymbol{w} \tag{4}$$

   From previous problem, we can get that every principal component $\boldsymbol{w}$ can be written as a linear combination of the data points, so:

$$\boldsymbol{w} = \boldsymbol{X} \cdot \boldsymbol{a} \tag{5}$$

   Replace $\boldsymbol{w}$ in equation (4) with equation (5), we can get the following fomulation:

$$\begin{aligned} &\max_{\boldsymbol{a}} \boldsymbol{a^T} \cdot \boldsymbol{X^T} (\boldsymbol{X} \cdot \boldsymbol{X}^T) \boldsymbol{X} \cdot \boldsymbol{a} \\ = &\max_{\boldsymbol{a}} \boldsymbol{a}^T \cdot (\boldsymbol{X}^T \boldsymbol{X} \cdot \boldsymbol{X}^T \boldsymbol{X}) \cdot \boldsymbol{a} \end{aligned} \tag{6}$$

   Because we can express the matrix $\boldsymbol{G}$ as:

$$\boldsymbol{G} = \boldsymbol{X}^T \cdot \boldsymbol{X} \tag{7}$$

   So the objective of maximizing variance between projected data points is:

$$\max_{\boldsymbol{a}} \boldsymbol{a}^T \cdot \boldsymbol{G}^2 \cdot \boldsymbol{a} \tag{8}$$

3. Solution:
   The variance of data points in the direction of principle component $\boldsymbol{w}$ is:

$$\boldsymbol{var} = \boldsymbol{w}^T(\boldsymbol{X} \cdot \boldsymbol{X}^T)\boldsymbol{w} \tag{9}$$

   Because $\boldsymbol{w}$ is unit vector, so:

$$\boldsymbol{w}^T \cdot \boldsymbol{w} = 1 \tag{10}$$

   Given equation (1) and (10), equation (10) can be written as:

$$\boldsymbol{var} = \boldsymbol{w}^T(\boldsymbol{X} \cdot \boldsymbol{X}^T)\boldsymbol{w} = \lambda \boldsymbol{w}^T \cdot \boldsymbol{w} = \lambda \tag{11}$$

   From problem 1, we know that:

$$\boldsymbol{a} = \frac{\boldsymbol{X}^T \cdot \boldsymbol{w}}{\lambda} \tag{12}$$

   where $\boldsymbol{X}^T \cdot \boldsymbol{w}$ denotes the vector of projections of data points onto the principal component $\boldsymbol{w}$. From equation (11), the variance of data points in that direction is λ.
   Obviously, $\boldsymbol{a}$ is the vector of projections of data points onto the principal component, normalized by the variance of data points in that direction.

4. Solution:
   Because the data is not centered, the mean of feature is:

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i} = \frac{1}{n} \cdot \boldsymbol{X} \cdot \boldsymbol{1} \tag{13}$$

   After centering, the new data would be:

$$\begin{aligned} \boldsymbol{X}' &= \boldsymbol{X} - \overline{\boldsymbol{X}} \\ &= \boldsymbol{X} - \frac{1}{n} \cdot \boldsymbol{X} \cdot \boldsymbol{1} \end{aligned} \tag{14}$$

   According to equation (4), (5), (6), we can get a new objective by replacing $\boldsymbol{X}$ in (6) with $\boldsymbol{X}'$. Equation (5) would be:

$$\boldsymbol{w} = \boldsymbol{X}' \cdot \boldsymbol{a} \tag{15}$$

   Therefore, the new objective is:

$$\max_{\boldsymbol{a}} \boldsymbol{a}^T \cdot (\boldsymbol{X}'^{\boldsymbol{T}}\boldsymbol{X}' \cdot \boldsymbol{X}'^{\boldsymbol{T}}\boldsymbol{X}') \cdot \boldsymbol{a} \tag{16}$$

From equation (14), we can get:

$$
\begin{aligned}
&\boldsymbol{X'^T} \cdot \boldsymbol{X'} \\
&\qquad = (\boldsymbol{X^T} - \frac{1}{n} \cdot \boldsymbol{1^T} \cdot \boldsymbol{X^T}) \cdot (\boldsymbol{X} - \frac{1}{n} \cdot \boldsymbol{X} \cdot \boldsymbol{1}) \\
&\qquad = \boldsymbol{X^T} \cdot \boldsymbol{X} - \frac{1}{n} \cdot \boldsymbol{1^T} \cdot \boldsymbol{X^T} \cdot \boldsymbol{X} \\
&\qquad - \frac{1}{n} \cdot \boldsymbol{X^T} \cdot \boldsymbol{X} \cdot \boldsymbol{1} + \frac{1}{n^2} \cdot \boldsymbol{1} \cdot \boldsymbol{X^T} \cdot \boldsymbol{X} \cdot \boldsymbol{1} \\
&\qquad = \boldsymbol{G} - \frac{1}{n} \cdot \boldsymbol{1^T} \cdot \boldsymbol{G} - \frac{1}{n} \cdot \boldsymbol{G} \cdot \boldsymbol{1} + \frac{1}{n^2} \cdot \boldsymbol{1} \cdot \boldsymbol{G} \cdot \boldsymbol{1}
\end{aligned}
\tag{17}
$$

So the final solution for objective is:

$$\max_{\boldsymbol{a}} \boldsymbol{a}^T \cdot \boldsymbol{G'^2} \cdot \boldsymbol{a} \tag{18}$$

where $G^{'}$ is $\boldsymbol{G} - \frac{1}{n} \cdot \boldsymbol{1^T} \cdot \boldsymbol{G} - \frac{1}{n} \cdot \boldsymbol{G} \cdot \boldsymbol{1} + \frac{1}{n^2} \cdot \boldsymbol{1} \cdot \boldsymbol{G} \cdot \boldsymbol{1}$.

3 SVM and Newton's Method

1. Task 3

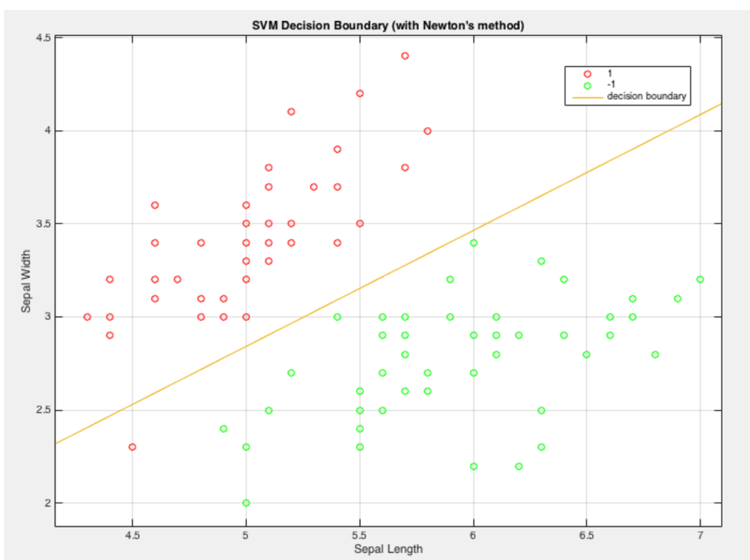

Figure 1: Result Visualization