CMSE 820 HW6

This HW is due on Oct 27th at 11:59 pm.

Question 1: For PCA, the main goals are

a. Encode training data (Reduce dimension, or finding $Y \in \mathbb{R}^{d\times n}$)

b. Reconstruct training data (Denoising)

c. Encode testing example (a new data $x^* \in \mathbb{R}^p$)

d. Reconstruct test example.

Given a $\mathbf{X} \in \mathbb{R}^{p\times n}$ data matrix, a new data point $x^* \in \mathbb{R}^p$ and a feature map $\phi : \mathbb{R}^p \to \mathbb{R}^\infty$, if we perform PCA on the feature space, whether can we still achieve the four goals? If so, how to ? If not, why?

Question 2: Implement the Kernel PCA (you can use any language) using the following two kernels

- Radial Basis (Gaussian) kernel (σ^2 is tuning parameter) :

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

- Polynomial Kernel (a is tuning parameter)

$$k(x, y) = (x^T y)^a.$$

  For this kernel, it is better to center the data first.

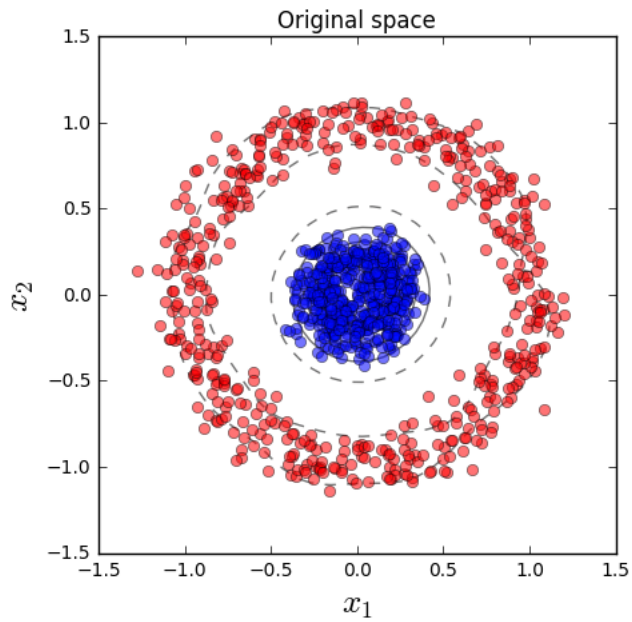
You are only allowed to use either SVD or Eigen decomposition functions from existing packages.

Question 3: Download "HW6dat.csv" from D2L. This file contains 1000 data points with 2 dimensions (x_1 and x_2 being its dimensions) and a label y.

(1) Make a plot similar to the following figure

(2) Run PCA on this dataset and make a scatering plot using PC1 and PC2.

(3) Run KPCA using two different kernels from Question 2 and compare the results with PCA. Explain your results including what casues these differences and which one are better?



(4) Project the following out-of-sample points: (0, -0.7), (0, 0.7), (0.7, 0), and (-0.7, 0) to first PC learn from PCA, the two Kernel principal components learned from KPCA with Radial Basis kernel and Polynomial Kernel. Plot your results.

Question 4: MDS of cities: Go to the following website geobytes.com Perform the following experiment.

(1) Input the following cities: Boston, Chicago, DC, Denver, LA, Miami, New York City, Seattle, and San Francisco, and collect the pairwise air traveling distances shown on the website in to a matrix D_1; Also,

(2) Run your own codes of classic MDS and plot the normalized eigenvalues $\lambda_i/(\sum_i \lambda_i)$ in a descending order of magnitudes, analyze your observations (did you see any negative eigenvalues? if yes, why?);

(3) Make a scatter plot of those cities using top 2 or 3 eigenvectors, and analyze your observations.

(4) Now, go to google map, and replace the distance matrix D with driving distance and repeat the analysis. Interpret your results.

(5) Now, go to google map, and replace the distance matrix D with driving hours and repeat the analysis. Interpret your results.

Question 5: *Positive Semi-definiteness:* Recall that a n-by-n real symmetric matrix K is called positive semi-definite (p.s.d. or $K \succeq 0$) iff for every $x \in \mathbb{R}^n, x^T K x \geq 0$. Assume $A \succeq 0$ and $B \succeq 0$ ($A, B \in \mathbb{R}^{n \times n}$).

(1) Show that $K \succeq 0$ if and only if its eigenvalues are all nonnegative.

(2) Show that $d_{ij} = K_{ii} + K_{jj} - 2K_{ij}$ is a squared distance function, i.e. there exists vectors $u_i, u_j \in \mathbb{R}^n$ for $1 \leq i, j \leq n$ such that $d_{ij} = \|u_i - u_j\|^2$.

(3) Show that $A + B \succeq 0$ (elementwise sum), and $A \circ B = \{A_{ij}B_{ij}\} \succeq 0$ (Hadamard product or elementwise product).

(4) Show that the eigen values of AB are all positive given A, B are symmetric positive definite matrix with the same dimension.

(5) If $A \succeq 0$ and $c \geq 0$, then $cA \succeq 0$

(6) If $A \succeq 0$ and C can be written as $C = \{t_{[i]}A_{ij}t_{[j]}\}$ for $\forall t \in \mathbb{R}^n$, then $C \succeq$.

(7) Show that the Hadamard integral power $A^{\circ p} = \{A_{ij}^p\}$ with $p \in \mathbb{N}$ and Hadamard exponential $\exp(\circ A) = \{\exp(A_{ij})\}$ are p.s.d..