CMSE 820 HW5

This HW is due on Oct 13th at 11:59 pm.

Question 1:

a. If $f_s(\cdot)$ is convex for any $s \in S$, prove that $f(x) = max_{s \in S} f_s(x)$ is convex. Note that the set S here (number of functions $f_s(\cdot)$) can be infinite.

b. Prove that the dual problem for any general minimization problem is a convex optimization problem.

Question 2: Consider quadratic program:

$$\min_x \frac{1}{2} x^T Q x + c^T x$$
$$\text{subject to } Ax = b, x \geq 0,$$

where $x \in \mathbb{R}^p$ and $Q \in \mathbb{R}^{p \times p}$ is a positive definite matrix. Now, derive its dual problem.

Question 3: For PCA, the main goals are

a. Encode training data (Reduce dimension)

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

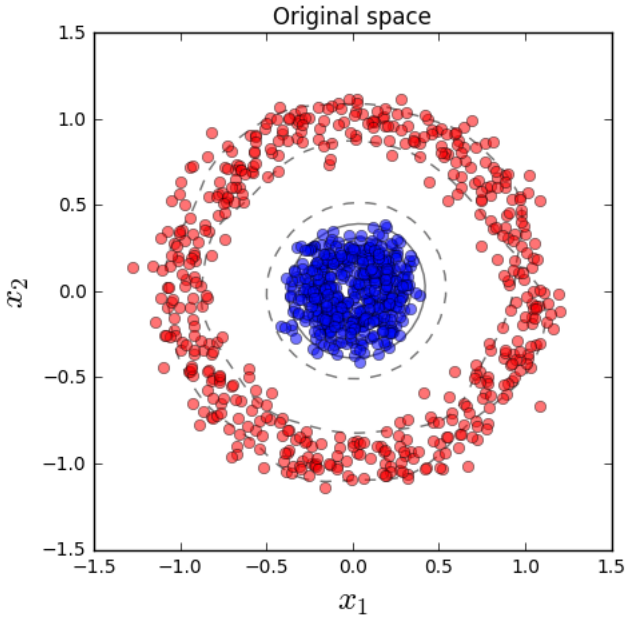
- Radial Basis (Gaussian) kernel (σ^2 is tuning parameter) :

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

- 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 5: Download "HW5dat.csv" from D2L. This file contains 1000 datapoints with 2 dimensions (x_1 and x_2 being its dimensions) and a lable y.

(1) Make a plot similar to

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

(3) Run KPCA using two different kernels 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.