Midterm, CMSC 828J – Spring 2013
Assigned, 3/25/13 – Due 4/3/2013

Instructions

Students must work individually to answer these questions. It is ok to discuss material covered in class with other students, but not to share ideas aimed at solving any particular questions on the exam. For example, it is all right to review Kernel PCA with other students, but not to discuss any thoughts on how this material applies to question 1. The exam is "open book". Students may consult any material, including material on the class web page. It is ok to look for additional reference material that might assist in solving these problems. Any material that you benefit from that is not on the class web page should be cited.

Please write up your exams neatly, and give complete and coherent answers to all questions. If we cannot follow your reasoning in solving a question with modest effort, your solution will be marked as incorrect.

Question 1 PCA, Kernel PCA

Let Φ be a mapping from a 2D space to a 5D feature space given as: $\Phi(x, y) = (x^2, y^2, xy, x, y)$. Is there a set of data points, (x_i, y_i) so that PCA in the original space will find that the magnitude of the two principal components of these points are bigger than 0, but kernel PCA in the feature space will find that only the first principal component has non-zero magnitude? If yes, give an example. If no, show why this is not possible.

Question 2 Wavelets, Fourier

Consider the function $f(x)$ defined for $x \in [0, 2\pi)$

$$f(x) = \begin{cases} 0 & : x < 1 \\ 1 & : x \geq 1 \end{cases} \tag{1}$$

(a) Write this as a linear combination of c (a constant function), $sin(kx)$ and $cos(kx)$.

(b) Write it as a linear combination of Haar wavelets (including a constant function).

(c) What does this tell us about the sparseness of images when represented in the wavelet domain.

Question 3 Sparse Projections

Suppose we have an unknown point, p, in a 3D space. We project it onto two known, randomly chosen directions, r_1 and r_2, so that we know $< p, r_1 >$ and $< p, r_2 >$.

(a) We now attempt to reconstruct our original point by finding a 3D point that produces these projections, and that has the smallest possible L_1 norm. True or false: this reconstructed point will always be the sparsest point that produces these two projections (ie the point with the smallest L_0 norm)? Either prove this is true or give an example where it is false.

(b) Suppose we reconstruct the point by minimizing the L_2 distance instead of the L_1 distance. True or false: this reconstructed point will never be the sparsest point that produces these projections? Either prove this is true or give an example where it is false.

Question 4 Multi-modal learning

In this question we will look into different optimization problems to learn multi-modal projection directions and try to analyze them. Notations for this question: $X_i \ \forall\ i \in \{1, 2\}$ are data matrices in the i^{th} view, with one column per data sample. The j^{th} column of X_1 and X_2 represent paired data in view 1 and 2, respectively. If the dimension of the data samples in the i^{th} view is d_i and there are a total of n samples, X_i will be $d_i \times n$.

We use MATLAB notation in representing matrix operations. If this is unclear, please ask.

Optimization 1

Let's make a matrix $Z = [X_1; X_2]$ (that is, X_1 is on top of X_2). Now let's solve this optimization problem

$$\begin{aligned} W, H = \text{argmin} \left\| Z - \hat{W}\hat{H} \right\|_F^2 \\ s.t. \quad \hat{W}^T\hat{W} = I_m \end{aligned} \tag{2}$$

Here, W will be a $(d_1 + d_2) \times m$ with $m \ll min(d_1, d_2)$ and H is $m \times n$ and I_m is the $m \times m$ Identity matrix.

- **Q1** - How would you solve this problem? Prove that your solution does in fact minimize (2).

- **Q2** - Let's denote $W_1 = W(1 : d_1, :)$ and $W_2 = W(d_1 + 1 : end, :)$. Consider a column of W_1 and the corresponding column of W_2. What is the relationship between these columns? **Hint:** We have provided data matrices X_1 and X_2 as face images of 200 person under two different poses in matlab format, in case you want to see what W_1 and W_2 look like (see class web page).

- **Q3** - Are $W_1^T W_1$ and $W_2^T W_2$ diagonal matrices?

- **Q4** We can use this model to generate a latent space for new paired data. For example, suppose that x_1^{new} and x_2^{new} are paired data, drawn from the same distribution as X_1 and X_2. Then we can find a latent space representation by solving: $W_1 h_1^{new} = x_1^{new}$ and $W_2 h_2^{new} = x_2^{new}$, and using h^{new} and h_2^{new} as the latent space representation of x_1^{new} and x_2^{new}. This latent space will be successful when, for paired data, $h_1^{new} \approx h_2^{new}$. Use first 100 paired-subjects supplied in the mat files to learn W_1 and W_2 with different values of $m \in \{10, 20, 30, 40, 50, 60\}$ and plot the ratio of $corr(h_1^i, h_2^i)$ and $corr(x_1^i, x_2^i)$ for the remaining 100 subject pairs for different values of m. If implemented properly, it will convince you that indeed the learned latent space brings correspnding points together in the latent space, which are otherwise far in the original representations.

  (a) Give an example of a simple, analytically described distribution for X_1 and X_2 for which this condition will hold?

  (b) Give an example of a simple, analytically described distribution for X_1 and X_2 for which this condition will not hold?

- **Q5** - Can you find a kernel version of this problem, such that, $X_i \to \Phi_i$ and $W_i \to \tilde{W}_i$. Here, Φ_i $(d_{H_i} \times n)$ and $\tilde{W}_i$ $(d_{H_i} \times m)$ are the matrices with data samples and directions mapped to a Hilbert space with d_{H_i} being the dimensionality of the Hilbert space, which can possibly be infinite as well. We are given the corresponding kernels for the Hilbert spaces as $K_1(\phi_1^i, \phi_1^j) = K_1^{ij}$ and $K_2(\phi_2^i, \phi_2^j) = K_2^{ij}$, for view 1 and 2 respectively. Show the steps of kernelization or state why it is not possible to kernelize this problem. (Hint: consider the answer to Q2 in solving this problem).

Optimization 2

Lets make a matrix $Z = X_1 X_2^T$. Now let's solve the following optimization problem

$$\begin{aligned} W, H = \text{argmin} & \left\| Z - \hat{W}\hat{H} \right\|_F^2 \\ s.t. \quad & \hat{W}^T\hat{W} = \hat{H}\hat{H}^T = D \end{aligned} \tag{3}$$

Here, W is $d_1 \times m$ matrix, H is $m \times d_2$ matrix and D is $m \times m$ diagonal matrix.

- **Q1** - How would you solve this problem? Prove that your solution does in fact minimize (3).
- **Q2** We can also use this model to generate a latent space for new paired data. Suppose that x_1^{new} and x_2^{new} are paired data. How do we find their latent space representation? **Hint:** We have provided data matrices X_1 and X_2 as face images of 200 person under two different poses in matlab format, in case you want to see what W and H look like (see class web page). Use first 100 paired-subjects supplied in the mat files to learn W and H with different values of $m \in \{10, 20, 30, 40, 50, 60\}$ and plot the ratio of $corr(h_1^i, h_2^i)$ and $corr(x_1^i, x_2^i)$ for the remaining 100 subject pairs for different values of m. Here, h_1^i and h_2^i will be the latent space reprsentation for the paired data samples x_1^i and x_2^i. If implemented properly, it will convince you that indeed the learned latent space bring correspnding points together in the latent space, which are otherwise far in the original representations.
- **Q3** - Can you find a kernel version of this problem, such that, $X_i \to \Phi_i$, $W \to \tilde{W}$ and $H \to \tilde{H}$. Here, Φ_i ($d_{H_i} \times n$), $\tilde{W}$ ($d_{H_1} \times m$), and $\tilde{H}$ ($m \times d_{H_2}$) are the matrices with data samples and directions mapped to a Hilbert space with d_{H_i} being the dimensionality of the Hilbert space, which can possibly be infinite as well. We are given the corresponding kernels for the Hilbert spaces as $K_1(\phi_1^i, \phi_1^j) = K_1^{ij}$ and $K_2(\phi_2^i, \phi_2^j) = K_2^{ij}$, for view 1 and 2 respectively. Show the steps of kernelization or state why it is not possible to kernelize this problem. (Hint: consider the answer to Q2 in solving this problem).

Question 5 Manifolds

(a) Consider a unit sphere parameterized by projecting each hemisphere orthogonally into a plane. For example, we define $x_\alpha : U_\alpha \subset R^2 \to M$ so that

U_α is the unit disk in the $x - y$ plane, and $x_\alpha(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$. Note that $x_\alpha(U_\alpha)$ covers the upper hemisphere of the sphere only, and does not include points on the equator. We can similarly define three additional projections so that we cover the entire sphere, but we will not make use of these other projections in this problem. Give the local Riemannian metric at each point in the region $U\alpha$.

(b) Consider a manifold for which, in one chart, $x_\alpha : U_\alpha \subset R^2 \to M$, with U_α given by the subset of the $x - y$ plane with $x > 0, y > 0$, and for which the Riemannian metric is:

$$\begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix}$$

What is the length of a straight line segment from $(1, 1)$ to $(2, 2)$? Note, here I mean the line segment is straight according to normal Euclidean distance; it is not the geodesic path from $(1, 1)$ to $(2, 2)$ on this manifold.

(c) CHALLENGE PROBLEM: For the manifold and Riemannian metric given in (b), what is the equation for the geodesic that passes through the point $(1, 1)$ in the direction $(1, 1)$?