Homework 5

Handed out: Monday, March 18, 2019

Due: Monday, April 08, 2019 11:55 pm

Notes:

- We *highly* encourage typed (Latex or Word) homework. Compile as single report containing solutions, derivations, figures, etc.
- Submit all files including report pdf, report source files (e.g. .tex or .docx files), data, figures produced by computer codes and programs files (e.g. .py or .m files) in a **.zip** folder. Programs should include a Readme file with instructions on how to run your computer programs.
- Zipped folder should be turned in on Sakai with the following naming scheme: **HW5_LastName_FirstName.zip**
- Collaboration is encouraged however all submitted reports, programs, figures, etc. should be an individual student's writeup. Direct copying could be considered cheating.
- Homework problems that simply provide computer outputs with no technical discussion, Algorithms, etc. will receive no credit.
- Software resources set can be downloaded from this link.

1 Factor analysis

A. One problem with mixture models is that they only use a single latent variable to generate the observations. One model that addresses this issue is the Factor analysis. FA is a low rank parameterization of an MVN. In this context,

   (a) Derive an expression for the number of independent parameters in the factor analysis model.

   (b) Show that the factor analysis model is invariant under rotations of the latent space coordinates.

B. We consider a data set of $D = 11$ variables and $N = 387$ cases describing various aspects of cars, such as the engine size, the number of cylinders, the miles per gallon (MPG), the price, etc. Fit a $L = 2$ dimensional factor analysis model. Plot the scores in $\mathbb{R}^2$ to visualize the results. To get a better understanding of the "meaning" of the latent factors, project unit vectors corresponding to each of the feature dimensions $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$, $\boldsymbol{e}_2 = (0, 1, \ldots, 0)$ etc into the low dimensional space. Represent using **biplot**. Provide your discussion.

   The required data set can be downloaded from this link.

2 PCA and KPCA

A. Consider a linear-Gaussian latent-variable model having a latent space distribution $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\mathbf{0}, \mathbf{I})$ and a conditional distribution for the observed variable $p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\mathbf{W}\boldsymbol{z} + \boldsymbol{\mu}, \boldsymbol{\Phi})$ where $\boldsymbol{\Phi}$ is an arbitrary symmetric, positive definite noise covariance matrix. Now suppose that we make a non-singular linear transformation of the data variables $\boldsymbol{z} \rightarrow \mathbf{A}\boldsymbol{x}$, where $\mathbf{A}$ is a $D \times D$ matrix. If $\boldsymbol{\mu}_{ML}$, $\mathbf{W}_{ML}$ and $\boldsymbol{\Phi}_{ML}$ represent the maximum likelihood solution corresponding to the original untransformed data, show that $\mathbf{A}\boldsymbol{\mu}_{ML}$, $\mathbf{A}\mathbf{W}_{ML}$ and $\mathbf{A}\boldsymbol{\Phi}_{ML}\mathbf{A}^T$ will represent the corresponding maximum likelihood solution for the transformed data set. Finally, show that the form of the model is preserved in two cases:

(a) $\mathbf{A}$ is a diagonal matrix and $\boldsymbol{\Phi}$ is a diagonal matrix. This corresponds to the case of factor analysis. The transformed $\boldsymbol{\Phi}$ remains diagonal, and hence factor analysis is covariant under component-wise re-scaling of the data variables

(b) $\mathbf{A}$ is orthogonal and $\boldsymbol{\Phi}$ is proportional to the unit matrix so that $\boldsymbol{\Phi} = \sigma^2\mathbf{I}$. This corresponds to probabilistic PCA. The transformed $\boldsymbol{\Phi}$ matrix remains proportional to the unit matrix, and hence probabilistic PCA is covariant under a rotation of the axes of data space, as is the case for conventional PCA

B. For the data set given at this link, Compute the first 8 kernel principal component basis functions. Use RBF kernel with $\sigma^2 = 0.1$

3 Independent Component Analysis

A. Suppose that two variables z_1 and z_2 are independent. Show that the covariance matrix between these variables is diagonal. Now consider two variables y_1 and y_2 in which $-1 \leq y_1 \leq 1$ and $y_2 = Y_1^2$. Write down the conditional distribution $p(y_2|y_1)$ and show that this is dependent on Y_1. Now show that the covariance matrix between these two variables is again diagonal. This counter-example shows that zero correlation is not a sufficient condition for independence.

B. For some noisy observations of a 4D signal, use ICA to reconstruct the signal. The data set can be found at this link.

4 LASSO

A. Recall that the LASSO objective is to minimize $RSS(\beta) + \lambda \sum_j |\beta_j|$ whereas the ridge regression objective is to minimize $RSS(\beta) + \lambda \sum_j ||\beta_j||_2^2$.

(a) **True or False:** LASSO solutions result in sparsity in the regression coefficients. Explain.

(b) **True or False:** Ridge Regression solutions result in sparsity in the regression coefficients. Explain

(c) **True of False:** As we increase λ in LASSO, we expect the number of variables in our solution to increase. Explain.

(d) **True or False:** It is possible to achieve the least-squares solution using a LASSO objective. Explain

(e) **True or False:** Given any two LASSO solutions corresponding to λ_1 amd λ_2 with $\lambda_2 > \lambda_1$ and the same support for these two solutions, it is possible to write out a closed-form expression for all solutions corresponding to λ with $\lambda_1 < \lambda < \lambda_2$. Explain.

B. Generate sparse signal $\boldsymbol{w}^*$ of size $D = 4096$, consisting of 160 randomly placed ± 1 spikes. Next generate a random design matrix $\mathbf{X}$ of size $N \times D$ where $N = 1024$. Finally, generate a noisy observation $\boldsymbol{y} = \mathbf{X}\boldsymbol{w}^* + \boldsymbol{\epsilon}$ where $\epsilon_i = \mathcal{N}\left(0, 0.01^2\right)$. Estimate $\boldsymbol{w}$ from $\boldsymbol{y}$ and $\mathbf{X}$. Estimate $\boldsymbol{w}$ by using LASSO using $\lambda = 0.1\lambda_{max}$. Plot the results and discuss.

5 Automatic relevance determination and compressed sensing

A. Automatic relevance determination is another method that results in sparse solution.

(a) Explain (with equations and diagram) how ARD works and why it results in sparse solution.

(b) How is ARD connected to the MAP estimate?

B. Suppose we have an image which is corrupted in some way, e.g., by having text or scratches sparsely superimposed on top of it. We might want to estimate the underlying clean image. This is called image inpainting. One can use similar techniques for image denoising. One way to address this problem is to use compressed sensing

Write a code for image denoising using compressed sensing. Use it for the image provided at this link.

References