Assignment 5

CMSC 726: Machine Learning
November 20th, 2018

Name:

1. Problem 23.4 from the textbook.

2. (Programming Assignment)

   (a) Let $\mathbf{z} = [z_1, z_2]^T \in \mathbb{R}^2$. We have a collection of samples $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$ where each $\mathbf{z}^{(i)}$ follows a Gaussian distribution $\mathcal{N}(0, \mathbf{I}_2)$. Moreover, we have the matrix $\mathbf{A} \in \mathbb{R}^{10\times 2}$ where A_{ij} follows a uniform $[0, 1]$ distribution. Each data point $\mathbf{z}^{(i)}$ generates $\mathbf{x}^{(i)}$ according to $\mathbf{x}^{(i)} = \mathbf{A}\mathbf{z}^{(i)} + \epsilon\mathbf{w}^{(i)}$ where $\epsilon = 0.01$ and $\mathbf{w}^{(i)}$ follows a Gaussian distribution $\mathcal{N}(0, \mathbf{I}_{10})$. Construct a data set consisting of $1,000$ points for $\mathbf{x}^{(i)}$ according to outlined generative process.

   (b) Implement PCA over the generated dataset. Use the top two PCs to compute low-dimensional representations of points, i.e. for $\mathbf{x}^{(i)}$ find $\hat{\mathbf{z}}^{(i)} = [\hat{z}_1^{(i)}, \hat{z}_2^{(i)}]^T$. Make two scatter plots (one for each component) comparing the original values $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)}$ and the estimated $\hat{\mathbf{z}}^{(1)}, \ldots, \hat{\mathbf{z}}^{(m)}$. What is the mean-squared error between the two sets? (i.e., $1/m \sum_i \|\mathbf{z}^{(i)} - \hat{\mathbf{z}}^{(i)}\|^2$).

      Note :You can use a function from a python toolbox to do the eigendecomposition or SVD.

   (c) Repeat (a)-(b) this for the case where $\mathbf{x}^{(i)} = \mathbf{A}\mathbf{z}^{(i)}{\mathbf{z}^{(i)}}^T\mathbf{b} + \epsilon\mathbf{w}^{(i)}$ where $\mathbf{b} \in \mathbb{R}^{2\times 1}$ and both elements in $\mathbf{b}$ are generated from a uniform $[0, 1]$.