ECE 445 (ML for ENGG): Homework #3

Waheed U. Bajwa (waheed.bajwa@rutgers.edu)

1. Describe in words the differences and similarities between "feature engineering" and "feature learning."

2. A group of medical professionals working within both US and Canada have collected healthcare data from their patients. These data range from weight (lbs for US and kg for Canada) and height (inches for US and cms for Canada) to temperature (Fahrenheit for US and Celsius for Canada). Can you suggest some preprocessing of these data before the data samples can be used for analytics?

3. A fitness tracker company has collected the following data from its user base comprising 1,000,000 users over a period of 24 hours: pulse rate collected from each user every 1.5 minutes. We collect these data samples into a single matrix $\mathbf{X}$, in which columns corresponds to the data of a single user.
   - What are the dimensions of $\mathbf{X}$?
   - What is the dimensionality of data corresponding to each user?
   - Suppose we transform the data matrix $\mathbf{X}$ into a feature matrix $\widetilde{\mathbf{X}}$ using principal components analysis (using top k principal components, with $k = 100$).
     - What is the dimension of $\widetilde{\mathbf{X}}$?
     - What is the dimension of feature vectors corresponding to each user?
     - Can we go back from the feature vector of a user to (an approximation) of its original data? If yes, then describe the mathematical operation. If no, then explain the reason for your answer.

4. Consider a data matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ with $N \gg M$, whose singular value decomposition is given by $\mathbf{X} = \mathbf{U\Sigma V}^T$. Let $\mathbf{P}^* \in \mathbb{R}^{k \times M}$ denote the projection matrix associated with the top k principal components of $\mathbf{X}$. Derive an expression for the singular value decomposition of the **projected data matrix** $\widehat{\mathbf{X}} = \mathbf{P}^*\mathbf{X}$. You will need to remind yourself of the relationship between the singular value decomposition of $\mathbf{X}$ and the eigen value decomposition of $\mathbf{XX}^T$ for this problem.

5. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix whose singular value decomposition is given by $\mathbf{A} = \mathbf{U\Sigma V}^T$. Then a linear algebra fact is that

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{\min(n,m)} \sigma_i^2},$$

   where σ_i denotes the i-th singular value of $\mathbf{A}$. Using this fact and your result from the previous problem, prove the following fact concerning the residual error in the projected data matrix $\widehat{\mathbf{X}}$:

$$\text{PCA Error} := \|\widehat{\mathbf{X}} - \mathbf{X}\|_F^2 = \sum_{i=k+1}^{M} \sigma_i^2.$$

   Note: This expression for the PCA error provides a way to select the "best" k for principal components analysis. Specifically, one can plot the singular values of the data matrix $\mathbf{X}$ and select k to be the number such that the residual error is less than a small percentage of the total "energy" in $\mathbf{X}$, measured using $\|\mathbf{X}\|_F^2$; say, $\frac{\text{PCA Error}}{\|\mathbf{X}\|_F^2} \leq 5\%$.

6. Suppose that a machine learning problem involves 1000-dimensional data samples $\{\mathbf{x}_i\}_{i=1}^N$. It is known that each data sample can be well approximated by the following mathematical expression:

$$\mathbf{x}_i = \mathbf{A}\mathbf{z}_i,$$

   where $\mathbf{A} \in \mathbb{R}^{1000 \times 125}$ is an unknown but fixed matrix and $\mathbf{z}_i \in \mathbb{R}^{125}$ is an unknown vector.

- What is the maximum k that one should use for principal component analysis of the data matrix $\mathbf{X}$ that comprises the data samples $\{\mathbf{x}_i\}_{i=1}^N$? Explain your reasoning.
- Using the maximum k derived in the last part, what is the relationship between the "column space" of $\mathbf{A}$ and the top k principal components of $\mathbf{X}$?