Practice Questions

1. With the SVD, we write $X = UDV^T$. For each of the following matrices, state whether its eigenvectors are equal to the columns of U. For each case, prove your claim.

   (a) $X^T X$

   (b) XX^T

   (c) $X^T XX^T X$

   (d) $XX^T XX^T$

2. Argue whether each of the following statements are true concerning PCA and SVD. Back up your argument with the corresponding mathematical reasoning. Here you can assume the data matrix is an $n \times p$ matrix X, where $n > p$ and X has rank r, $r < p$.

   (a) The principal component directions are eigenvectors of the centered data matrix.

   (b) The principal component directions are right singular vectors of the centered data matrix.

   (c) The principal component scores are eigenvectors of the sample covariance matrix.

   (d) The sample variance of the $(r+2)^{nd}$ principal component scores $(\xi_{1(r+2)}, ..., \xi_{n(r+2)})$ is zero.

3. Show that the inner product of the vectors containing the projections of the centered data X^C onto the directions of the j^{th} and k^{th} principal component directions is always zero if $j \neq k$. In other words, prove

$$\sum_{i=1}^{n} \xi_{ij}\xi_{ik} = 0.$$

4. Suppose you have a dataset with $n = 100$ observations and $p = 10$ features.

   (a) Give an example of a principal component vector ϕ_1 (in terms of its elements) which can be interpreted as a weighted average of all 10 of the features.

   (b) Sketch a plot which depicts the elements of ϕ_1 across the feature index

   (c) Give an example of a principal component vector ϕ_2 (in terms of its elements) which can be interpreted as a contrast between the subset of features 1-5 and the subset of features 6-10.

   (d) Sketch a plot which depicts the elements of ϕ_2 across the feature index

   (e) Calculate the inner product of ϕ_1 and ϕ_2.

5. A local bookstore has a very loyal user base and a database of rating (1-5). They send monthly emails with personalized book recommendations. Their recommendations are based on a version of matrix factorization in which the user embedding matrix U has been already computed (with some previous data). They do monthly updates of the item matrix V. In this setting consider the following questions:

   (a) How many parameters get updated in gradient descent each month?

   (b) Consider matrices U, V_1, V_2 and the following test set. Test set: $y_{0,0} = 1$, $y_{5,3} = 5$ Which matrix V_1 or V_2 fits better the test data? (Show your work.)

$$U = \begin{bmatrix} 0.2 & 2.4 & -0.1 \\ 1.6 & 1 & 2 \\ 7 & -6 & 1.5 \\ 5 & 0.7 & 2 \\ 2 & 0.4 & 3 \\ -1 & 1 & 0 \\ 0.8 & -0.9 & 0.9 \end{bmatrix}, V_1 = \begin{bmatrix} 0 & 1 & 1 \\ -3 & 0 & 0.1 \\ 0 & 4 & -2 \\ 0 & 1 & 3 \end{bmatrix}, V_2 = \begin{bmatrix} 0 & 0 & -10 \\ 6 & 0 & 0.1 \\ 0 & 4 & 9 \\ -3 & 1 & 0.2 \end{bmatrix}$$

6. You have a binary classification problem with two features, $\mathbf{x}_i = (x_{i1}, x_{i2}) \in \mathbb{R}^2,\ i = 1, ..., n$. Consider a logistic regression model with an interaction term given below.

$$\hat{y} = \sigma(w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_1 \cdot x_2 + b)$$

   (a) What are the parameters in this equation?

   (b) First assume you have only one observation and compute the gradient descent equations for the parameters using the log loss function as loss:

$$L(y, \hat{y}) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})).$$

   (c) Given a training set $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, write one of the equations from part (b) using the whole dataset.

Hints: Feel free to use σ' as the derivative for σ. The derivative of $\log x$ is $\frac{1}{x}$.

7. Derive the Gradient Descent updating equations for an arbitrary element $u_{\ell k}$ of the U matrix in the Matrix Factorization model for collaborative filtering.