CS/ECE/ME532 Activity 12

Estimated time: 5 mins for Q1, 15 mins for Q2 (review), 20 mins for Q3, and 20 mins for Q4.

1. Let the n-by-p rank-r ($n > p > r$) matrix $\boldsymbol{X}$ have SVD $\boldsymbol{X} = \boldsymbol{U\Sigma V}^T$ where $\boldsymbol{U}$ is n-by-r, $\boldsymbol{\Sigma}$ is r-by-r, and $\boldsymbol{V}$ is p-by-r.

a) Find the SVD of $\boldsymbol{Z} = \boldsymbol{X}^T$ in terms of $\boldsymbol{U}, \boldsymbol{\Sigma}$, and $\boldsymbol{V}$.

b) Find the orthonormal basis for the best rank-1 subspace to approximate the rows of $\boldsymbol{Z}$ in terms of $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{\Sigma}$.

2. Uniqueness of solutions and Tikhonov regularization (ridge regression). The least-squares problem is $\min_{\boldsymbol{w}} ||\boldsymbol{y} - \boldsymbol{Xw}||_2^2$. Assume $\boldsymbol{X}$ is n-by-p with $p < n$.

a) Under what conditions is the solution to the least-squares problem not unique?

b) The Tikhonov-regularized least-squares problem is

$$\min_{\boldsymbol{w}} ||\boldsymbol{y} - \boldsymbol{Xw}||_2^2 + \lambda ||\boldsymbol{w}||_2^2$$

Show that this can be written as an ordinary least-squares problem $\min_{\boldsymbol{w}} ||\hat{\boldsymbol{y}} - \hat{\boldsymbol{X}}\boldsymbol{w}||_2^2$ and find $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{X}}$.

c) Use the results from the previous part to determine the conditions for which the Tikhonov-regularized least-squares problem has a unique solution.

3. Psuedoinverse and truncated SVD. The solution to the ridge regression problem

$$\min_{\boldsymbol{w}} ||\boldsymbol{y} - \boldsymbol{Xw}||_2^2 + \lambda ||\boldsymbol{w}||_2^2$$

is given by $\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. The *psuedoinverse* of $\boldsymbol{X}$, denoted $\boldsymbol{X}^\dagger$, can be defined by looking at the limit of the ridge regression solution as $\lambda \to 0$ (from above):

$$\boldsymbol{X}^\dagger = \lim_{\lambda \downarrow 0} (\boldsymbol{X}^T\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}^T.$$

a) Let $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, $p \leq n$, have SVD $\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{V}^T = \sum_{i=1}^p \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$. Show that

$$(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T = \sum_{i=1}^{p} \frac{\sigma_i}{\sigma_i^2 + \lambda}\boldsymbol{v}_i\boldsymbol{u}_i^T.$$

Hint: Note that $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}\Sigma^2\boldsymbol{V}^T$ *and* $\lambda\boldsymbol{I} = \boldsymbol{V}\lambda\boldsymbol{I}\boldsymbol{V}^T$.

b) Using the limit definition of the psuedoinverse above, show that when $\boldsymbol{X}^T\boldsymbol{X}$ is invertible, then $\boldsymbol{X}^\dagger = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$.

c) Argue that when $\boldsymbol{X}$ is square and invertible, then $\boldsymbol{X}^\dagger = \boldsymbol{X}^{-1}$.

d) Argue that if $\boldsymbol{X}$ is rank $r < p$, then for $\lambda > 0$,

$$(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T = \sum_{i=1}^{r} \frac{\sigma_i}{\sigma_i^2 + \lambda}\boldsymbol{v}_i\boldsymbol{u}_i^T.$$

e) Now argue that if $\boldsymbol{X}$ is rank $r < p$,

$$\boldsymbol{X}^\dagger = \sum_{i=1}^{r} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T = \boldsymbol{V}\Sigma_r^{-1}\boldsymbol{U}^T$$

where Σ_r^{-1} is a matrix with $1/\sigma_i$ on the diagonal for $i = 1, \ldots, r$, and zero elsewhere.

4. The data file is available with a matrix $\boldsymbol{X}$ of 100 three-dimensional data points. A script is available with code to assist you with visualizing and fitting this data. Use the results of the SVD to find $\boldsymbol{a}$, a basis for the best (minimum sum of squared distances) one-dimensional subspace for the data.

a) Run the code to display the data in Figure the first figure. Use the rotate tool to inspect the scatter plot from different angles. Does the data appear to lie very close to a one-dimensional subspace? Does the data appear to be zero mean?

b) Figure 2 depicts the centered data and the one-dimensional subspace that contains the dominant feature you identified using the SVD. Use the rotate tool to inspect the data and one-dimensional subspace from different angles. Is a one-dimensional subspace a reasonable fit to the data? Comment on the error.

c) Now comment out (insert %) the line of code that subtracts the mean of the data. Does the dominant feature identified by SVD continue to be a good fit to the data? Comment on the importance of removing the mean before performing PCA.