CS/ECE/ME 532

Unit 4 Practice Problems

1. In two dimensions, PCA with one component is equivalent to linear regression for fitting a line to data points. Explain why or why not.

SOLUTION: False. While PCA approximates the data with a line, the line it chooses minimizes the perpendicular distance from the line to the data. Regression fits a line that minimizes the vertical distance from the line to the data.

2. Consider the low-rank approximation to a matrix $\boldsymbol{A} \approx \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$. The term bias refers to $\sum_{i=1}^{r} \sigma_i$.

SOLUTION: False. Bias is the part of the matrix that is missing, which has norm $\sum_{i=r+1}^{M} \sigma_i$.

3. What algorithm is used for ranking pages on the internet. Select all that apply.

a) PageRank

b) Iterative Singular Value Thresholding

c) K-means

d) Power iterations

e) Spectral methods

SOLUTION: PageRank and Power iterations

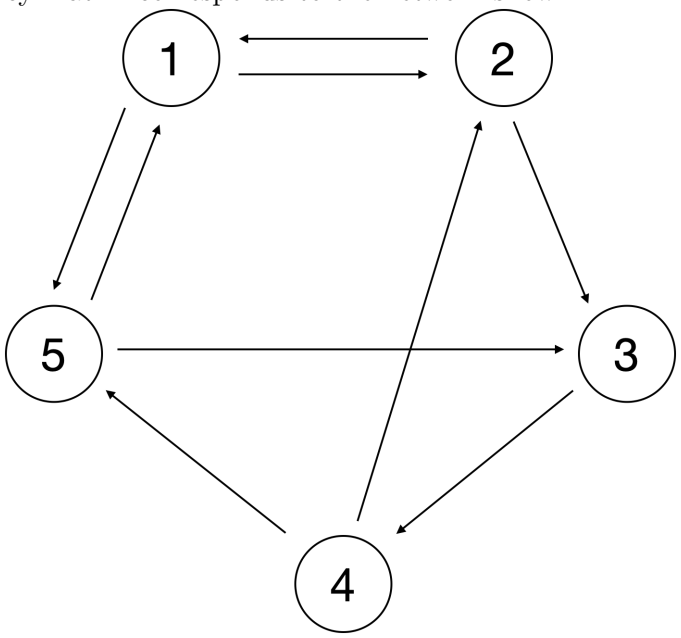
4. Suppose $\boldsymbol{A}$ is a square, symmetric, positive definite matrix and has eigendecomposition $\boldsymbol{A} = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}^T$ and singular value decomposition $\boldsymbol{A} = \sum_{i=1}^{M} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$. What is the relationship between the largest eigenvalue and the singular value decomposition?

SOLUTION: The eigendecomposition and singular value decomposition for square, positive definite, symmetric matrices are identical if the eigenvalues are ordered from largest to smallest, so the largest eigenvalue is equal to σ_1.

5. You measure 500 three-dimensional vectors $\boldsymbol{x}_i, i = 1, 2, \ldots 500$. Explain how to find the best fitting line to the data.

SOLUTION: First find the mean (average) $\boldsymbol{x}_m = \frac{1}{500} \sum_{i=1}^{5} 00 \boldsymbol{x}_i$ and remove it $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \boldsymbol{x}_m$. Then put the mean removed data in a matrix $\tilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{\boldsymbol{x}}_1 & \tilde{\boldsymbol{x}}_2 & \cdots & \tilde{\boldsymbol{x}}_5 00 \end{bmatrix}$ and find the SVD $\tilde{\boldsymbol{X}} = \sum_{i=1}^{3} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$. The best fitting line to the mean removed data is given in terms of the first principal component as $\alpha \boldsymbol{u}_1$, so the best line to the original data is defined as $\boldsymbol{x}_m + \alpha \boldsymbol{u}_1$.

6. Which normalized adjacency matrix corresponds to the network shown?



a) $\mathbf{A} = \begin{bmatrix} 0 & 0 & 1/2 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1/3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 \end{bmatrix}$

b) $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$

c) $\mathbf{A} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 1 \\ 0 & 0 & 1/2 & 1 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \end{bmatrix}$

d) $\mathbf{A} = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$

SOLUTION: Choice d)