6CCS3AIN, Tutorial 05 (Version 1.0)

1. You're given the data matrix $\mathbf{X}$ with $n = 4$ houses and $d = 3$ features. Your goal is to run the PCA algorithm from the slides to $\mathbf{X}$ in order to reduce the number of dimensions to $k = 2$.

$$\mathbf{X} = \begin{pmatrix} 4 & 2 & 3 \\ 6 & 1 & 3 \\ 4 & 2 & 5 \\ 7 & 8 & 3 \end{pmatrix}$$

6CCS3AIN, Tutorial 05 Answers (Version 1.0)

1. The exact numbers are calculate in the `PCA_tutorial_example.py`. We will use approximations here. Our data matrix is.

$$\mathbf{X} = \begin{pmatrix} 4 & 2 & 3 \\ 6 & 1 & 3 \\ 4 & 2 & 5 \\ 7 & 8 & 3 \end{pmatrix}$$

   (a) Step 1: Compute the mean row vector

$$\bar{\mathbf{x}}^T = \begin{pmatrix} 5.25 & 3.25 & 3.5 \end{pmatrix}$$

   (b) Step 2: Compute the mean row matrix

$$\bar{\mathbf{X}} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \bar{\mathbf{x}}^T = \begin{pmatrix} 5.25 & 3.25 & 3.5 \\ 5.25 & 3.25 & 3.5 \\ 5.25 & 3.25 & 3.5 \\ 5.25 & 3.25 & 3.5 \end{pmatrix}$$

   (c) Step 3: Subtract mean (obtain mean centred data)

$$B = \mathbf{X} - \bar{\mathbf{X}} = \begin{pmatrix} -1.25 & -1.25 & -0.5 \\ 0.75 & -2.25 & -0.5 \\ -1.25 & -1.25 & 1.5 \\ 1.75 & 4.75 & -0.5 \end{pmatrix}$$

   The dimensions are $n \times d$

   (d) Step 4: Compute the covariance matrix of rows of B

$$C = \frac{1}{n} B^T B = \begin{pmatrix} 1.68 & 2.44 & -0.63 \\ 2.43 & 7.69 & -0.63 \\ -0.63 & -0.63 & 0.75 \end{pmatrix},$$

   where $n = 4$ The dimensions are $(d \times n) \times (n \times d) = d \times d$

   (e) Step 5: Compute the k largest eigenvectors $\mathbf{v}_1, \mathbf{v}_1, \dots, \mathbf{v}_k$ (not covered how to do this in this module. You use Python or WolframAlpha).
   Each eigenvector has dimensions $1 \times d$

$$\mathbf{v}_1 \approx \begin{pmatrix} -0.34 \\ -0.94 \\ 0.1 \end{pmatrix}$$

$$\mathbf{v}_2 \approx \begin{pmatrix} -0.71 \\ 0.32 \\ 0.63 \end{pmatrix}$$

   (f) Step 6: Compute matrix $\mathbf{W}$ of k-largest eigenvectors

$$\mathbf{W} = \begin{pmatrix} -0.34 & -0.71 \\ -0.94 & 0.32 \\ 0.1 & 0.63 \end{pmatrix}$$

   Dimensions of $\mathbf{W}$ are $(d \times k)$.

   (g) Step 7: Multiply each datapoint $\mathbf{x}_i$ for $i \in \{1, 2, \dots, n\}$ with $\mathbf{W}^T$

$$\mathbf{W}^T = \begin{pmatrix} -0.34 & -0.94 & 0.1 \\ -0.71 & 0.32 & 0.63 \end{pmatrix}$$

   We have e.g.,

$$\mathbf{x}_1 = \begin{pmatrix} 4 \\ 2 \\ 3 \end{pmatrix}$$

$$\mathbf{y}_i = \mathbf{W}^T \cdot \mathbf{x}_i$$

So

$$\mathbf{y}_1 \approx \begin{pmatrix} -2.92 \\ -0.3 \end{pmatrix}$$

$$\mathbf{y}_2 \approx \begin{pmatrix} -2.66 \\ -2.03 \end{pmatrix}$$

$$\mathbf{y}_3 \approx \begin{pmatrix} -2.72 \\ 0.97 \end{pmatrix}$$

$$\mathbf{y}_4 \approx \begin{pmatrix} -9.55 \\ -0.48 \end{pmatrix}$$

Dimensions of $\mathbf{y}_i$ are $(k \times d) \times (d \times 1) = k \times 1$