2/4/25

Date __/__/__
Page

Lab-11

Principle component analysis

input:

- dataset $D = \{x_1, x_2, \ldots x_N\}$ where each $x_i \in R^d$
- no. of principle components (target dimensions(k)) where $k \leq d$

output

- Reduced dataset $Z = \{z_1, z_2, \ldots z_N\}$
- eigen values and eigen vectors

Algorithm

1. Standardize the data
   - compute the mean of each feature $j = 1, 2, \ldots d$

$$\mu_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij}$$

   - SD compute

$$\sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_{ij} - \mu_j)^2}$$

Date __/__/__
Page _____

- standardize each feature

$$x_{i,j} \leftarrow \frac{x_{i,j} - \mu_j}{\sigma_j}$$

2 compute covariance matrix

$$\Sigma = \frac{1}{N-1} X^T X$$

3. Compute the eigenvalues $\lambda_1, \ldots \lambda_n$ and eigen vectors $v_1, v_2, \ldots v_d$

sort the eigen values in descending order and select top K eigenvector

4. $W = [v_1, v_2, \ldots v_K]$

5 $Z = X W$

6 Return Z and W