Principle component Analysis

1 Overview

- **Dimension reduction**: model data $x \in R^d$ as "approximately" lying in some n-dimensional subspace, where $n \ll d$. The reason is that many components of x are dependent.

2 Project in one direction

- Dataset: suppose $S = \{x_1, x_2, \cdots, x_m\}$ contains m samples
- Centering the data by removing the mean $\bar{x} = \frac{1}{m}\sum_{i=1,2,\cdots,m} x_i$, i.e.,

$$x_i \leftarrow x_i - \bar{x} \quad \forall i \in [m]$$

- Assume the direction that we would like to project the data is u, then the objective function is to maximize the average of square of norms for all data samples, that is

$$J(u) = \frac{1}{m}\sum_{i=1}^{m}(u^\top x_i)^2$$

$$= u^\top\left(\frac{1}{m}\sum_{i=1}^{m} x_i x_i^\top\right)u$$

$$= u^\top \Sigma u$$

Where $||u||^2 = 1$.

- The optimization of finding a single projection direction is

$$\max_u J(u) = u^\top \Sigma u$$

$$\text{s.t.} \quad u^\top u = 1$$

- By the lagrange multiplier, the optimal solution satisfies

$$\Sigma u - \lambda u = 0$$

$$\Sigma u = \lambda u$$

Where λ is the **largest eigenvalue** and u is the **corresponding eigenvector**.

3 PCA

- Eigen decomposition of matrix Σ

$$U^{\top}\Sigma U = \Lambda$$

- Eigenvalues $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$
- Eigenvectors $U = [u_1, \ldots, u_d]$

- Reduce x_i to a n dimensional vector $y_i = V^T x_i$, choose the first n eigenvectors corresponding to the n largest eigenvalues among all d eigenvalues. Thus we have

$$V = [u_1, \cdots, u_n]$$

- The value n can be determined based on the accuracy of approximation, i.e.,

$$\frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq T$$

Where T is the threshold.

References:

Yangfeng Ji, machine learning, lecture notes

https://aman.ai/cs229/pca/