Lecture 7: Dimensionality Reduction

Course Teacher: Md. Shariful Islam Bhuyan

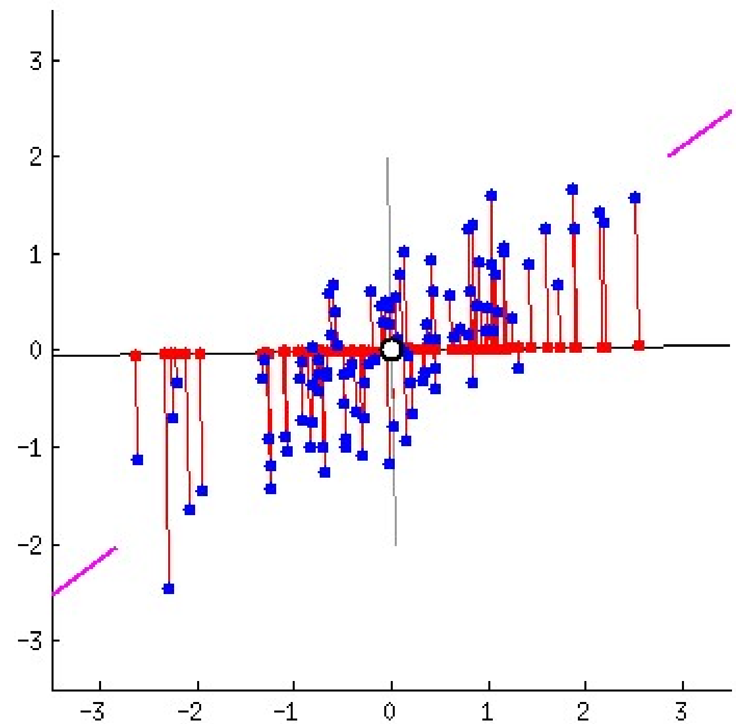
Linear Dimensionality Reduction

- Reduce dimensionality … number of feature K<<D
- $\mathbf{z}_{K\times 1} = \mathbf{U}_{K\times D} \times \mathbf{x}_{D\times 1}$
- Useful linear projection … matrix is orthogonal unit vectors … dot products give corresponding component.
- Defining useful
  - Capture variance, differentiating properties
  - Minimize reconstruction loss

Principal Component Analysis

- Full PCA is a rotation operation, change of axis
- First principal component (PC) finds direction of highest variance
- Second PC finds direction of next highest variability orthogonal to first PC
- Take top K PCs and project the data along those
- Maximizing variance and minimizing reconstruction loss is equivalent due to Pythagoras theorem

Projection



First Principal Component

- Projection $\mathbf{z}_n = \mathbf{u}_1^T \mathbf{x}_n$

$$\frac{1}{N}\left(\sum_{n=1}^{N} \mathbf{u}_1^T \mathbf{x}_n\right) = \mathbf{u}_1^T \boldsymbol{\mu}$$

- Variance

$$\frac{1}{N}\left(\sum_{n=1}^{N} (\mathbf{u}_1^T \mathbf{x}_n - \mathbf{u}_1^T \boldsymbol{\mu})^2\right) = \frac{1}{N}\left(\sum_{n=1}^{N} \{\mathbf{u}_1^T (\mathbf{x}_n - \boldsymbol{\mu})\}^2\right) = \frac{1}{N}\left(\sum_{n=1}^{N} \mathbf{u}_1^T (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T \mathbf{u}_1\right)$$

$$= \mathbf{u}_1^T \frac{1}{N}\left(\sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T\right) \mathbf{u}_1 = \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$$

- First principal component $= \arg\max_{\mathbf{u}_1} \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$

Eigen Decomposition

- To prevent trivial solution let $\mathbf{u_1}$ is a unit vector. Assume $\|\mathbf{u}_1\| = 1 = \mathbf{u}_1^T\mathbf{u}_1$
- Using Lagrange multiplier first principal component

$$= \arg\max_{\mathbf{u}_1} \mathbf{u}_1^T\mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^T\mathbf{u}_1)$$

$$\frac{\partial}{\partial \mathbf{u}_1}\{\mathbf{u}_1^T\mathbf{S}\mathbf{u}_1 + \lambda_1(1 - \mathbf{u}_1^T\mathbf{u}_1)\} = 2\mathbf{S}\mathbf{u}_1 - 2\lambda_1\mathbf{u}_1 = 0 \quad or,$$

$$\mathbf{S}\mathbf{u}_1 = \lambda_1\mathbf{u}_1 \left\{\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x}\right\}$$

- Which eigenvector? Where λ_1 is largest since $\mathbf{u}_1^T\mathbf{S}\mathbf{u}_1 = \mathbf{u}_1^T\lambda_1\mathbf{u}_1 = \lambda_1$
- Data ⟶ Covariance matrix ⟶ Eigen-decomposition ⟶ Top K PC

Further explanation

- https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues
- http://www.onmyphd.com/?p=lagrange.multipliers