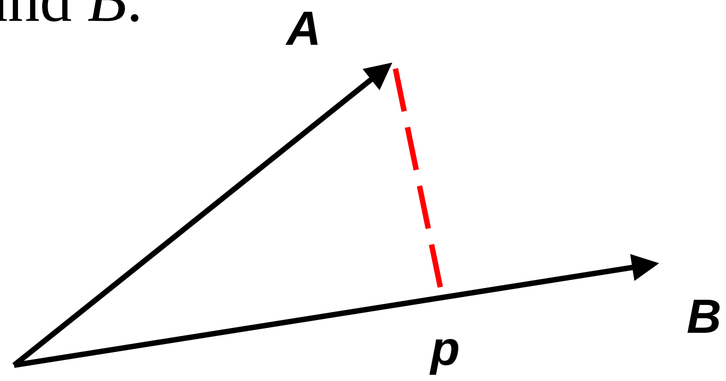
STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 6
September 12, 2019

Linear Projection

- Two vectors A and B.



- The projection of A on the line of B is computed as

$$p = A^T \frac{B}{\| B \|}$$

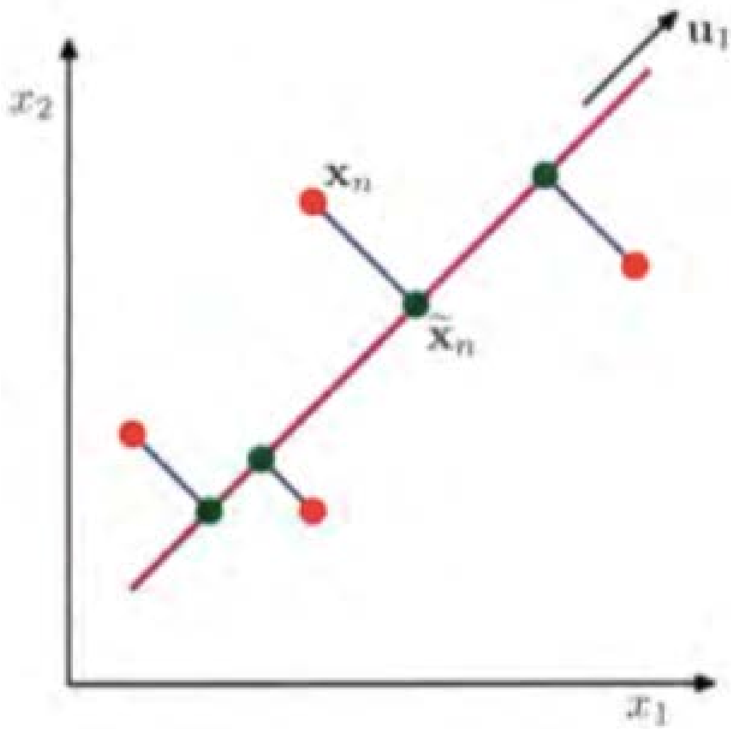
- When B is a unit vector (i.e. $\|B\|=1$), the projected value is

$$p = A^T B = B^T A.$$

Goal of PCA

- PCA seeks a space of lower dimensionality, known as the principal subspace and denoted by the magenta line, such that the orthogonal projection of the data points (red dots) onto the subspace maximize the variance of the projected points (green dots).

- Alternatively, we can minimize the sum-of-squares of the projection errors, indicated (blue lines).



One-Dimensional Projection

- Given x_1, …, x_N in $\mathbf{R}^D$, our goal is to project the data onto a 1-d space while mazximizing the variance of the projected data.

- Let the projection vector be u_1, then the projected value is

$$u_1^T x_n, \qquad n = 1, \ldots, N$$

- Let the sample mean be $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$. Then the variance of the projected data is given by

$$\frac{1}{N}\sum_{n=1}^{N}(u_1^T x_n - u_1^T \bar{x})^2 = u_1^T S u_1$$

where S is data covariance

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T.$$

Maximization

- Note that project vector is a unit vector, i.e. $\|u_1\| = 1$.
- We maximize the projected varaince with respect to u_1.
- Using Lagrange multiplier, we can maximize

$$u_1^T S u_1 + \lambda(1 - u_1^T u_1)$$

- By taking derivative with respect to u_1, we will have a stationary point when

$$S u_1 = \lambda u_1.$$

which says that u_1 must be an eigenvector of S.

Maximization

- Moreover, we see that the variance is given by

$$u_1^T S u_1 = \lambda.$$

- So the variance will be a maximum when we set $\boldsymbol{u}_1$ equal to the eigenvector having the **largest eigenvalue**.

- This **eigenvector** is known as the **first principal component**.

Multi-dimensional Projection

- We can define additional principal components in an incremental fashion by choosing each new direction to be that which maximizes the projected variance amongst all possible directions orthogonal to those already considered.

- We consider the general case of an M-dimensional projection space.

- **The optimal linear projection for which the variance of the projected data is now defined by the M eigenvectors $u_1, \ldots, u_M$ of the data covariance matrix S corresponding to the M largest eigenvalues $\lambda_1, \ldots, \lambda_M$.**

Projection

- Let the projection matrix be $V = [v_1,\dots,v_M,\dots, v_D]$. This is a orthogonal matrix in $\mathbf{R}^{D\mathrm{x}D}$, i.e. $VV^T = V^TV = I_D$.
- The submatrix $B = [v_1,\dots,v_M]$ with $B^TB = I_M$. Then the projected value is
$$B^T x_n, \qquad n = 1,\dots, N$$
- The total variance of the projected value is
$$\frac{1}{N}\sum_{n=1}^{N}(B^T x_n - B^T\bar{x})^T(B^T x_n - B^T\bar{x}) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})^T BB^T(x_n - \bar{x})$$
$$= \frac{1}{N}\sum_{n=1}^{N} tr\left((x_n - \bar{x})^T BB^T(x_n - \bar{x})\right) = \frac{1}{N}\sum_{n=1}^{N} tr\left(BB^T(x_n - \bar{x})(x_n - \bar{x})^T\right)$$
$$= tr\left(BB^T\frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T\right) = tr(BB^T\Sigma)$$

Projection

- Using singular value decomposition $\Sigma = U\Lambda U^T$, we have

$$tr(BB^T\Sigma) = tr(BB^TU\Lambda U^T) = tr(B^TU\Lambda U^TB)$$

- Denote the matrix U^TB using **column vectors**, i.e.

$$U^TB = [l_1, l_2, \ldots, l_D]^T.$$

- Note that $\Lambda = \text{diag}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_D)$. Then

$$tr(B^TU\Lambda U^TB) = tr\left([l_1, \ldots, l_D]\Lambda[l_1, \ldots, l_D]^T\right) = \sum_{i=1}^{D} \lambda_i l_i^T l_i$$

- We compute the sum

$$\sum_{i=1}^{D} l_i^T l_i = tr(B^TUU^TB) = tr(B^TB) = tr(I_M) = M$$

Projection

- Therefore, $\sum_{i=1}^{D} \lambda_i l_i^T l_i$ is a weighted sum of *D* eigenvalues, where all the weights are nonnegative and the sum is *M*.
- As the eigenvalues are in decreasing order, the optimal solution is

$$l_i^T l_i = \begin{cases} 1 & i = 1\ldots, M \\ 0 & i = M+1,\ldots, D \end{cases}$$

- We can re-write

$$U^T B = [l_1, l_2, \ldots, l_D]^T = [C, 0]^T$$

where *C* is in $\mathbf{R}^{MxM}$.

- Furthermore,

$$CC^T = [C,0][C,0]^T = (B^T U)(U^T B) = B^T B = I_M$$

Projection

- Therefore, C is orthogonal and

$$B = U[C, 0]^T = [u_1, \ldots, u_M] C^T.$$

- The projected values are

$$B^T x_n = C[u_1, \ldots, u_M]^T x_n = C[u_1^T x_n, \ldots, u_M^T x_n]^T.$$

- C is an orthogonal transformation within the projected subspace.

- We have proven that

The optimal linear projection for which the variance of the projected data is maximized is now defined by the M eigenvectors $u_1, \ldots, u_M$ of the data covariance matrix S corresponding to the M largest eigenvalues $\lambda_1, \ldots, \lambda_M$.