PCA sketch of the proof

CSCI 347

Adiesha Liyanage

Let the data D consist of n points over d attributes, i.e., it is an n × d matrix, given as

$$D = \begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_d \\ \hline x_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ x_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \cdots & \cdots & \ddots & \cdots \\ x_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array}$$

$$x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T \; spanned \; by \; d \; standard \; basis \; vectors$$

$$e_1, e_2, e_3, \dots, e_d$$

Standard basis vectors are **orthonormal**.

$$e_i^T e_j = 0 \;\; and \; \|e_i\| = 1$$

If we are given any other set of d **orthonormal** vectors $u_1, u_2, \cdots, u_d$ with $u_i^T u_j = 0 \ \ and \ \|u_i\| = 1$

We can re-express each point x as the linear combination

$$x = a_1 u_1 + a_2 u_2 + \cdots + a_d u_d$$

$a = (a_1, a_2, \ldots, a_d)^T$ *represents the coordinates of the* x *in the new basis.*

The above combination can be represented as a matrix multiplication as well.

$$x = Ua$$

Where U is the $d \times d$ matrix, whose i-th column comprises the i-the basis vector u_i

$$U = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{pmatrix}$$

Where U is an orthogonal matrix, where columns are orthonormal

$$U = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{pmatrix}$$

$$u_i^T u_j = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$$

Since U is an orthogonal matrix,

$$U^{-1} = U^T$$

$$U^{-1} = U^T \rightarrow U^T U = I$$

$$x = Ua$$
$$U^T x = U^T U a$$
$$U^T x = a$$
$$a = U^T x$$

Let's look at an example.

$$x = (-0.343, -0.754, 0.241)^T$$

$$e_1 = (1, 0, 0)^T, e_2 = (0, 1, 0)^T, e_3 = (0, 0, 1)^T$$

$$u_i^T u_j = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$$

New basis vectors

$$u_1 = (-0.390, 0.089, -0.916)^T, u_2 = (-0.639, -0.742, 0.200)^T, u_3 = (-0.663, 0.664, 0.346)^T$$

$$\|u_1\| = \sqrt{0.39^2 + 0.089^2 + 0.916^2} \approx 1$$

$$u_1^T u_3 = -0.39 * -0.663 + 0.089 * 0.664 - 0.916 * 0.346 \approx 0$$

$$a = U^T x = \begin{pmatrix} -0.39 & 0.089 & -0.916 \\ -0.639 & -0.742 & 0.2 \\ -0.663 & 0.664 & 0.346 \end{pmatrix} \begin{pmatrix} -0.343 \\ -0.754 \\ 0.241 \end{pmatrix} = \begin{pmatrix} -0.154 \\ 0.828 \\ -0.190 \end{pmatrix}$$

$$\boldsymbol{x = -0.15u_1 + 0.828u_2 - 0.19u_3}$$

MONTANA STATE UNIVERSITY

There are potentially infinite choices for the set of orthonormal basis vectors.

Example: Think about 3-dimension space. Take the 3-basis vector and rotate it in the space.

Let's look at projection of x into r basis vectors.

$$x = a_1 u_1 + a_2 u_2 + \cdots + \ a_d u_d$$
$$We\ pick\ first\ r\ basis\ vectors\ from\ this.$$
$$x' = a_1 u_1 + a_2 u_2 + \cdots + \ a_r u_r \quad , r \ll d$$

Note that x and x' are not the same.

Here x' is the projection of x onto the first r basis vectors.

We can write this in matrix notation as well.

$$x' = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = U_r a_r$$

U_r is the matrix comprising of the first r basis vectors, and a_r is the vector comprising the first r coordinates.

Remember:

$$a = U^T x$$

Restricting it to first r basis vectors,

$$a_r = U_r^T x$$

$$x' = \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_r \\ | & | & & | \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \end{pmatrix} = U_r a_r$$

$$a_r = U_r^T x$$

$$x' = U_r a_r = U_r U_r^T x$$

$$P_r = U_r U_r^T \text{ is the orthogonal projection matrix for the subspace spanned by the first } r \text{ basis vectors.}$$

$$x' = U_r a_r = U_r U_r^T x = P_r x$$

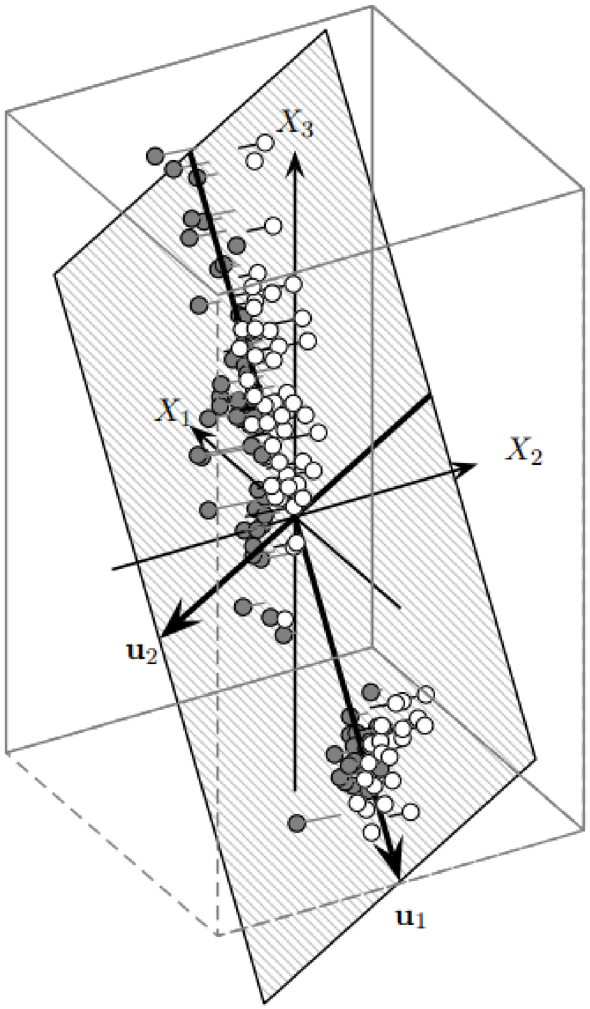
P_r is **symmetric** and **idempotent**, i.e.,

$$P_r = P_r^T \text{ and } P_r^2 = P_r$$

$$P_r = U_r U_r^T = \sum_{i=1}^{r} u_i u_i^T$$

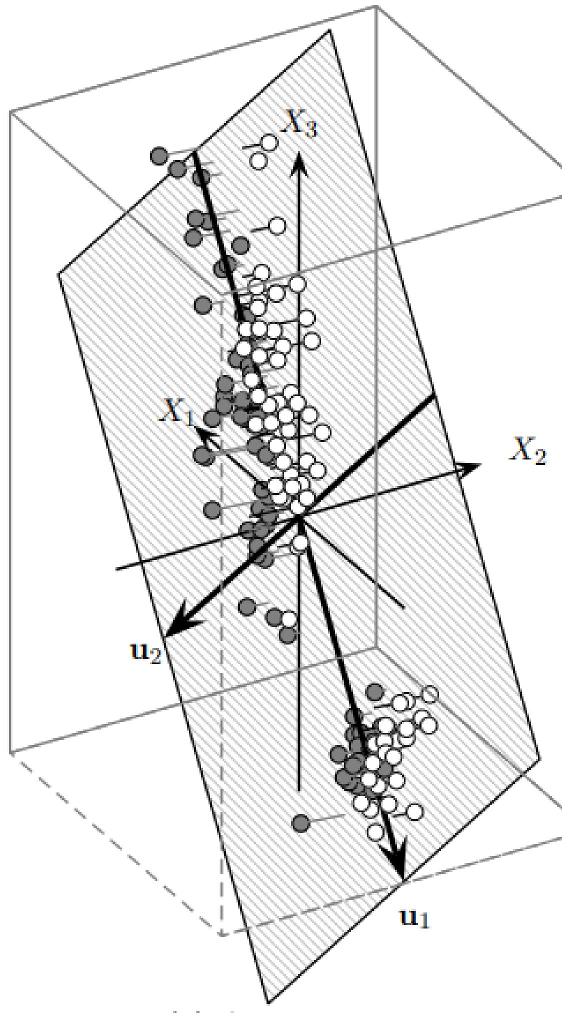
$$\epsilon = \sum_{i=r+1}^{d} a_i u_i = x - x'$$

We can get an expression for the error of the projection.



$$\epsilon = \sum_{i=r+1}^{d} a_i u_i = x - x'$$

We can get an expression for the error of the projection.



Note that x' and ϵ are orthogonal vectors.

MONTANA STATE UNIVERSITY

Let's look at an example.

$$x = (-0.343, -0.754, 0.241)^T$$

$$e_1 = (1, 0, 0)^T, e_2 = (0, 1, 0)^T, e_3 = (0, 0, 1)^T$$

New basis vectors

$$u_1 = (-0.390, 0.089, -0.916)^T, u_2 = (-0.639, -0.742, 0.200)^T, u_3 = (-0.663, 0.664, 0.346)^T$$

Let's pick only one new basis u_1 vector and project the data.

$$P_1 = u_1 u_1^T = \begin{pmatrix} -0.39 \\ 0.089 \\ -0.916 \end{pmatrix} \begin{pmatrix} -0.39 & 0.089 & -0.916 \end{pmatrix}$$
$$= \begin{pmatrix} 0.152 & -0.035 & 0.357 \\ -0.035 & 0.008 & -0.082 \\ 0.357 & -0.082 & 0.839 \end{pmatrix}$$

$$\boxed{\boldsymbol{x' = P_1 x = \begin{pmatrix} 0.060 \\ -0.014 \\ 0.141 \end{pmatrix}}}$$

$$\boxed{\boldsymbol{\epsilon = x - x' = a_2 u_2 + a_3 u_3 = \begin{pmatrix} -0.40 \\ -0.74 \\ 0.10 \end{pmatrix}}}$$

Let’s look at an example.

$$\boldsymbol{x'} = \boldsymbol{P_1 x} = \begin{pmatrix} \mathbf{0.060} \\ \mathbf{-0.014} \\ \mathbf{0.141} \end{pmatrix}$$

$$\boldsymbol{\epsilon} = \boldsymbol{x} - \boldsymbol{x'} = \boldsymbol{a_2 u_2} + \boldsymbol{a_3 u_3} = \begin{pmatrix} \mathbf{-0.40} \\ \mathbf{-0.74} \\ \mathbf{0.10} \end{pmatrix}$$

We can check that x' and ϵ are orthogonal.

$$\boldsymbol{x'^T e} = \begin{pmatrix} \mathbf{0.06} & \mathbf{-0.014} & \mathbf{0.141} \end{pmatrix} \begin{pmatrix} \mathbf{-0.40} \\ \mathbf{-0.74} \\ \mathbf{0.10} \end{pmatrix} = \mathbf{0}$$

Proof of PCA

- Assume that the data is centered.
- Let's think about projecting data onto one line.
- Projection of x_i on the vector u is:

$$x_i' = \left(\frac{u^T x_i}{u^T u}\right) u = (u^T x_i) u = a_i u$$

- Let's get an expression for the variance of the projected data points.
- $\sigma_u^2 = \frac{1}{n} \sum_{i=1}^{n} (a_i - \mu_u)^2$
- $\sigma_u^2 = \frac{1}{n} \sum_{i=1}^{n} (a_i)^2 \quad since\ \mu_u = 0$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu_u)^2$$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(u^T x_i)^2$$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(u^T x_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(u^T x_i)(u^T x_i) = \frac{1}{n}\sum_{i=1}^{n}(u^T x_i)\left(x_i^T u\right)$$

$$\sigma_u^2 = u^T\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right)u$$

$$(using\ associative\ and\ distributive\ properties)$$

$$\sigma_u^2 = u^T \Sigma u$$

We want to maximize σ_u^2 and find u that achieves that objective.

$$\sigma_u^2 = u^T \Sigma u$$

We want to maximize σ_u^2 and find u that achieves that objective while keeping the constraints $u^T u = 1$.

We can use Lagrangian multiplier.

$$\max_{a} J(u) = u^T \Sigma u - \alpha(u^T u - 1)$$

We take the derivative of $J(u)$ with respect to u and set it to zero.

$$\frac{\partial J(u)}{\partial x} = 0$$

$$2\Sigma u - 2\alpha u = 0$$

$$\Sigma u = \alpha u$$

This implies that α is an eignenvalue of the covariance matrix Σ.

$$\sigma_u^2 = u^T \Sigma u = u^T \alpha u = \alpha u^T u = \alpha$$

This means to maximize the variance choose the largest eigenvalue.