Lecture 17

Unsupervised Learning

Supervised problem: I have both X and Y and some training data to learn the rel. btwn X and Y

Un-supervised problem: I only have an X

— find/summarize patterns in the data

Ex. (1) dimensionality reduction:

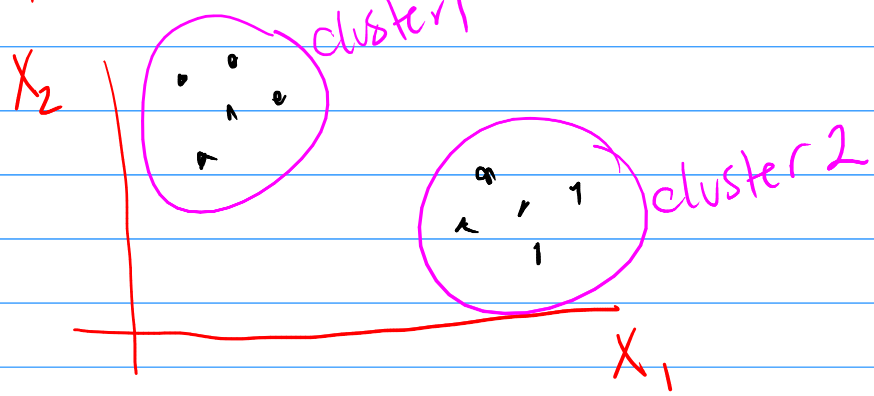
represent the data using fewer variables

p covariates $\rightsquigarrow$ q covariates

$q \ll p$

(2) clustering:

group data into similar "clusters"



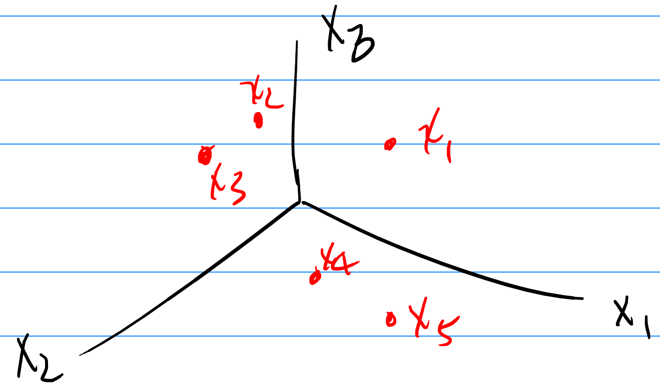
Principal Components Analysis (PCA)

Technique for dimension reduction.

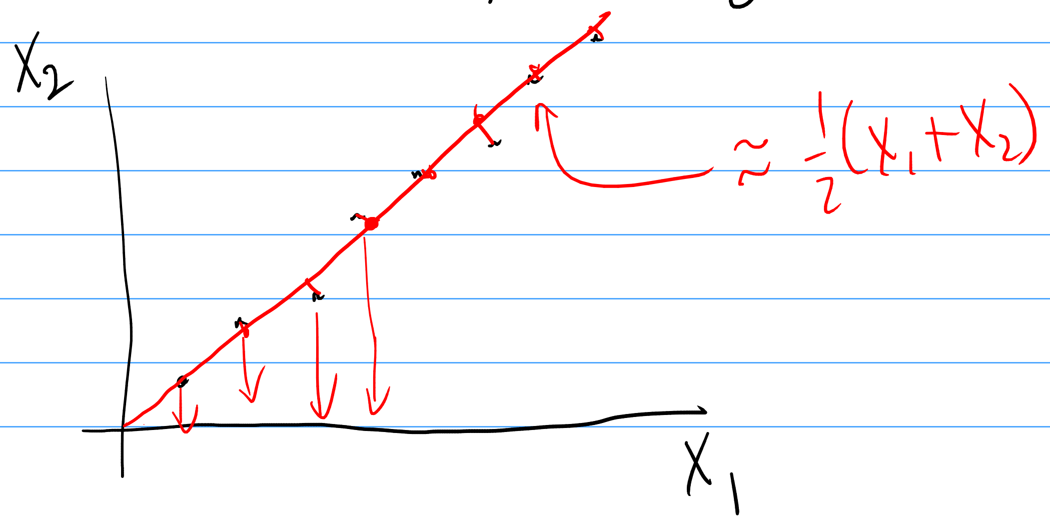
unsupervised setting:

$$X_{N \times P} = \begin{bmatrix} \text{— obs 1 —} \\ \text{— obs 2 —} \\ \vdots \end{bmatrix} = \begin{bmatrix} | & | & \\ \text{Var 1} & \text{Var 2} & \cdots \\ | & | & \end{bmatrix}$$

Visualize: $X = \begin{bmatrix} \text{—} x_1 \text{—} \\ \vdots \\ \text{—} x_N \text{—} \end{bmatrix}$ when $x_n \in \mathbb{R}^P$



Dim Reduction: can I get away w/ fewer dims w/o losing too much.



Goals of PCA

(1) reduce the # vars

$$X_1, \ldots, X_p \rightsquigarrow Z_1, Z_2, \ldots, Z_q$$

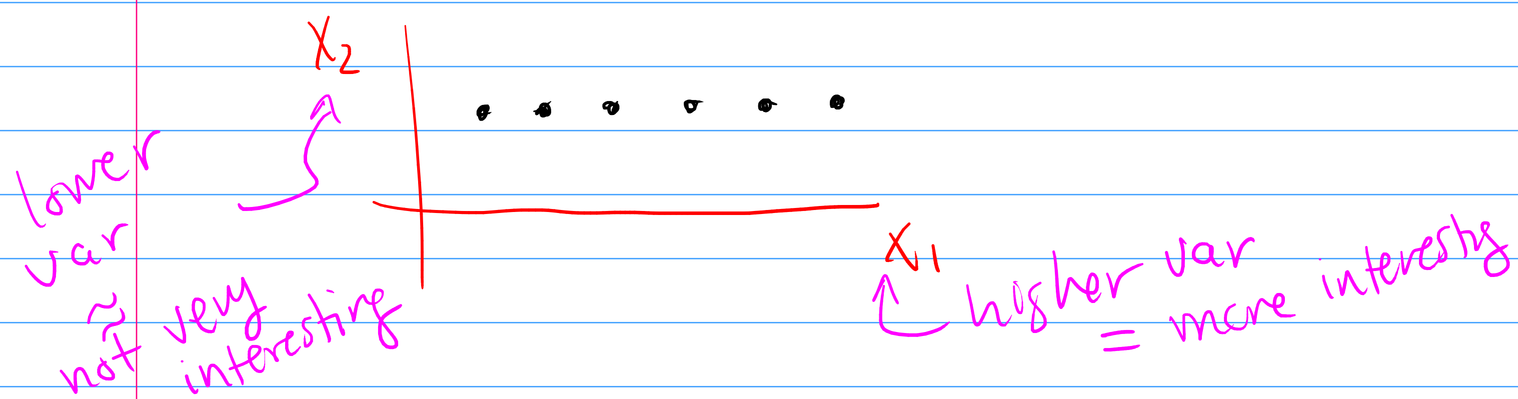
$$q \ll p$$

(2) don't lose too info. when doing this.

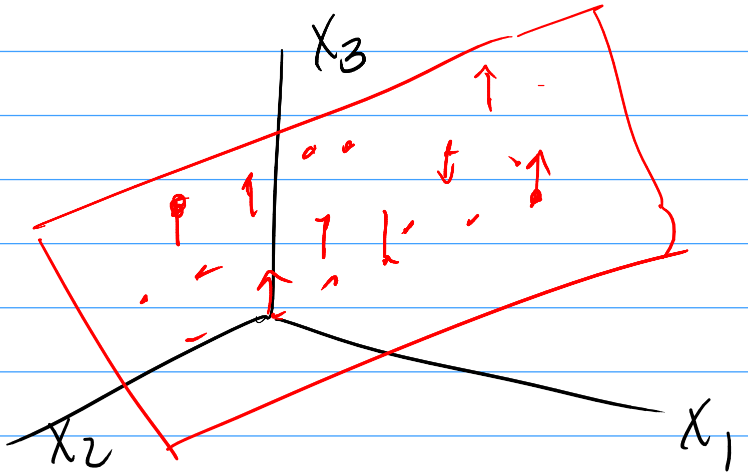
$\Updownarrow$

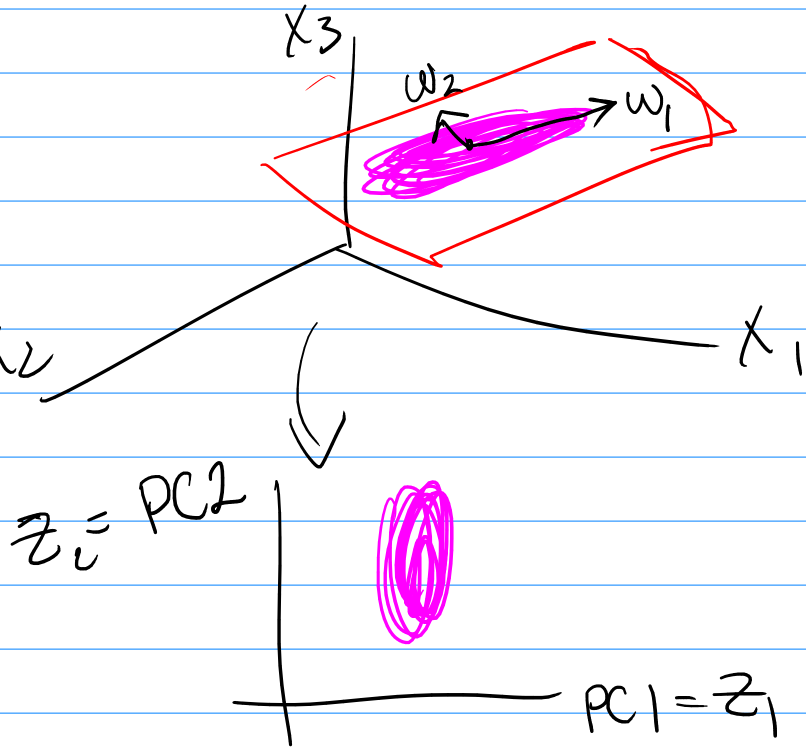
maximize the amount of info retained

<u>central dogma of PCA</u>: variance = info.



PCA operates by projecting data onto a lower-dimensional subspace.





Need to find basis of this subspace

$w_1, w_2, \ldots, w_q$

$z_i = Xw_i = w_{i1}X_1 + w_{i2}X_2 + w_{i3}X_3 + \cdots w_{ip}X_p$

ith coordinate in this subspace

z_i = LC of X_is

PCA: Find $\overbrace{\text{LCs of my } X\text{s}}^{z\text{s}}$ that maximize the variance of Zs subject to the zs being uncorrelated and the w_is are unit vectors.

Mathematically

$$\underset{N \times q}{Z} = XW$$

$W = \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_q \\ | & | & & | \end{bmatrix}$

i.e. $z_i = Xw_i$

Want to find W to

(1) maximize total variance of Z

$$\text{Tot Var}(Z) = \sum_{i=1}^{q} \hat{\text{Var}}(z_i)$$

Subject to

(1) $\text{Cor}(z_i, z_j) = 0$ (2) w_i to be unit vectors

Review: $x = (x_1, \dots, x_N)$ then

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{N} 1^T x$$

$(1, 1, \dots, 1)$

If $\bar{x} = 0$ (in PCA we're going to mean-center our vars)

$$Var(x) = \frac{1}{N-1}\sum_n (x_n - \bar{x})^2 = \frac{1}{N-1}\sum_n x_n^2 = \frac{1}{N-1}\|x\|^2$$

$$\propto \|x\|^2$$

If $\bar{x} = 0$ or $\bar{y} = 0$

$$\propto x^T x$$

$$Cov(x, y) \propto x^T y$$

So $Cov(x,y) = 0 \Leftrightarrow Cor(x,y) = 0$

$\Leftrightarrow x^T y = 0$

$\Leftrightarrow$ orthogonality

Consider

$$A = W^T X^T X W$$

(*) Assume cols of X have mean zero

$$A_{ij} = w_i^T X^T X w_j$$

$$= (Xw_i)^T (Xw_j)$$

$$= z_i^T z_j \propto cor(z_i, z_j)$$

PCA: find W to max $tr(A)$ subject to A is diagonal (cols W are unit vecs)

$$A_{ii} \propto Var(z_i)$$