PRINCIPAL COMPONENTS ANALYSIS

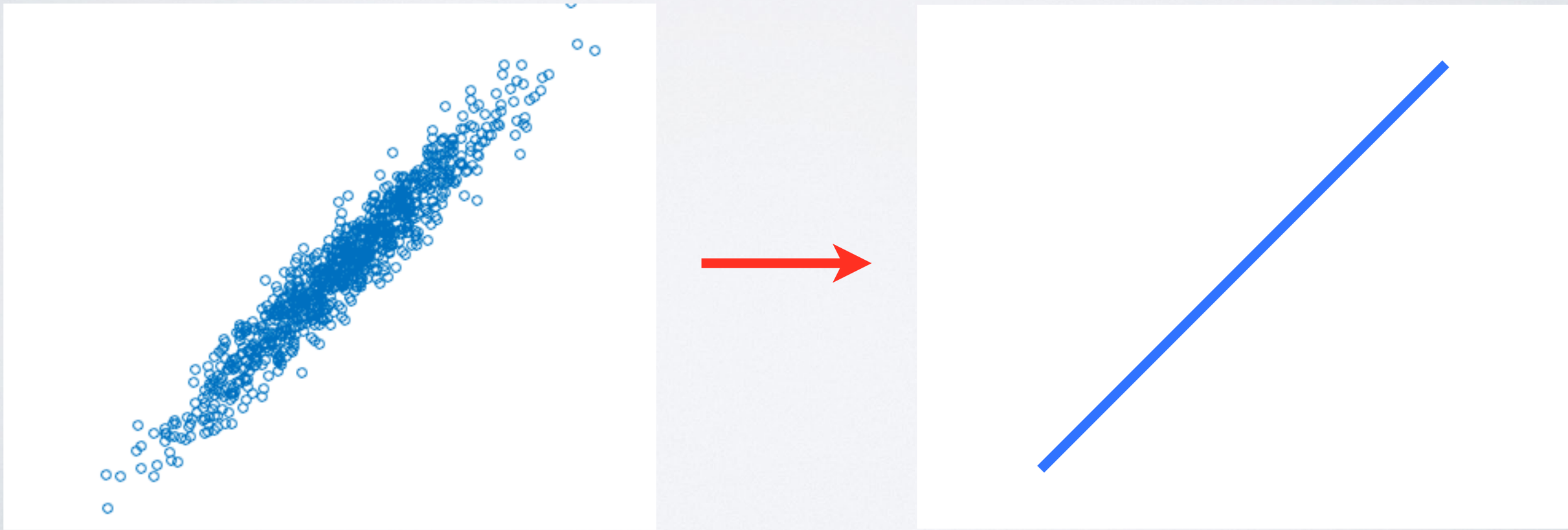
Machine Learning HS18

PCA APPLICATIONS

- Dimensionality Reduction
- Data Visualization
- Data Classification
- Noise Reduction
- etc.

INTRODUCTION

- PCA identifies the low dimensional subspace in which the data approximately lies



EXAMPLE 1

- Suppose we are given a dataset of attributes of m different types of automobiles:

$$\left\{x^{(i)}; i = 1,\ldots,m\right\} \quad x^{(i)} \in \mathbb{R}^n \text{ for each } i(n \ll m)$$

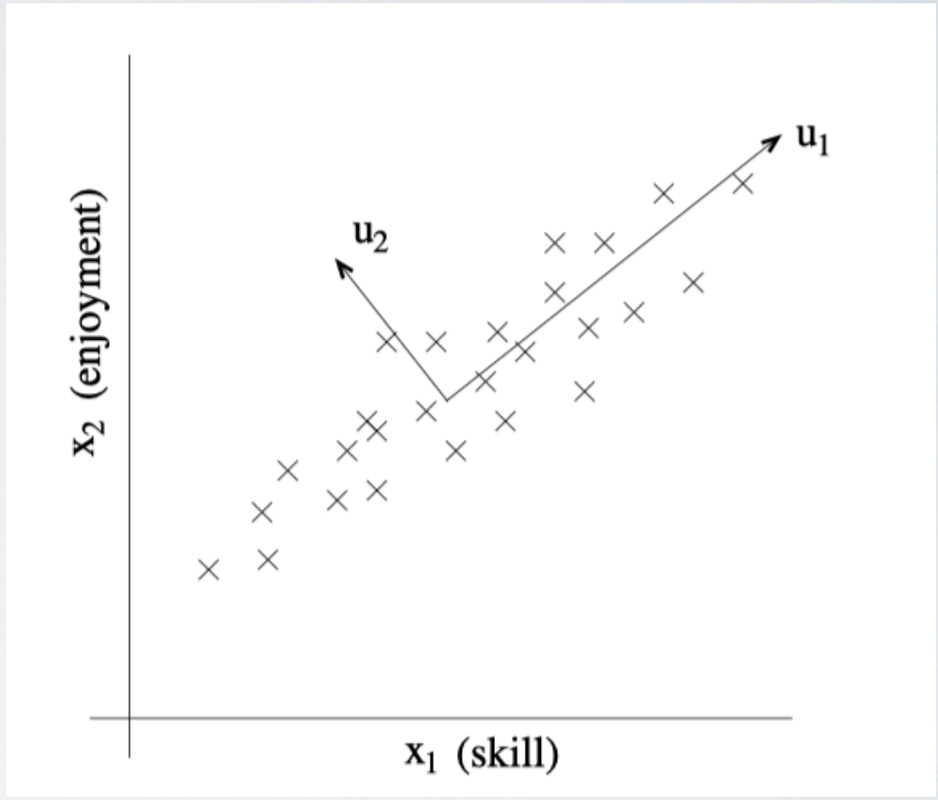
- Suppose two different attributes give a cars maximum speed measured in kilometers and miles per hour respectively

Observation: attributes are linearly dependent and therefore redundant

Conclusion: In reality data lies on $n-1$ dimensional subspace

EXAMPLE II

- Consider the dataset resulting from a survey of pilots for a radio-controlled helicopters
- Only the most committed students become good pilots
- Two attributes are very correlated
- We might assume that the data actually lies along some diagonal axis capturing the correlation
- Notice that this axis is the major axis of variation



GOAL: Project the data onto the vector that retains as much as possible of the data variation.

How can we automatically compute this direction?

PRE-PROCESSING THE DATA

Let $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$

Replace each $x^{(i)}$ with $x^{(i)} - \mu$

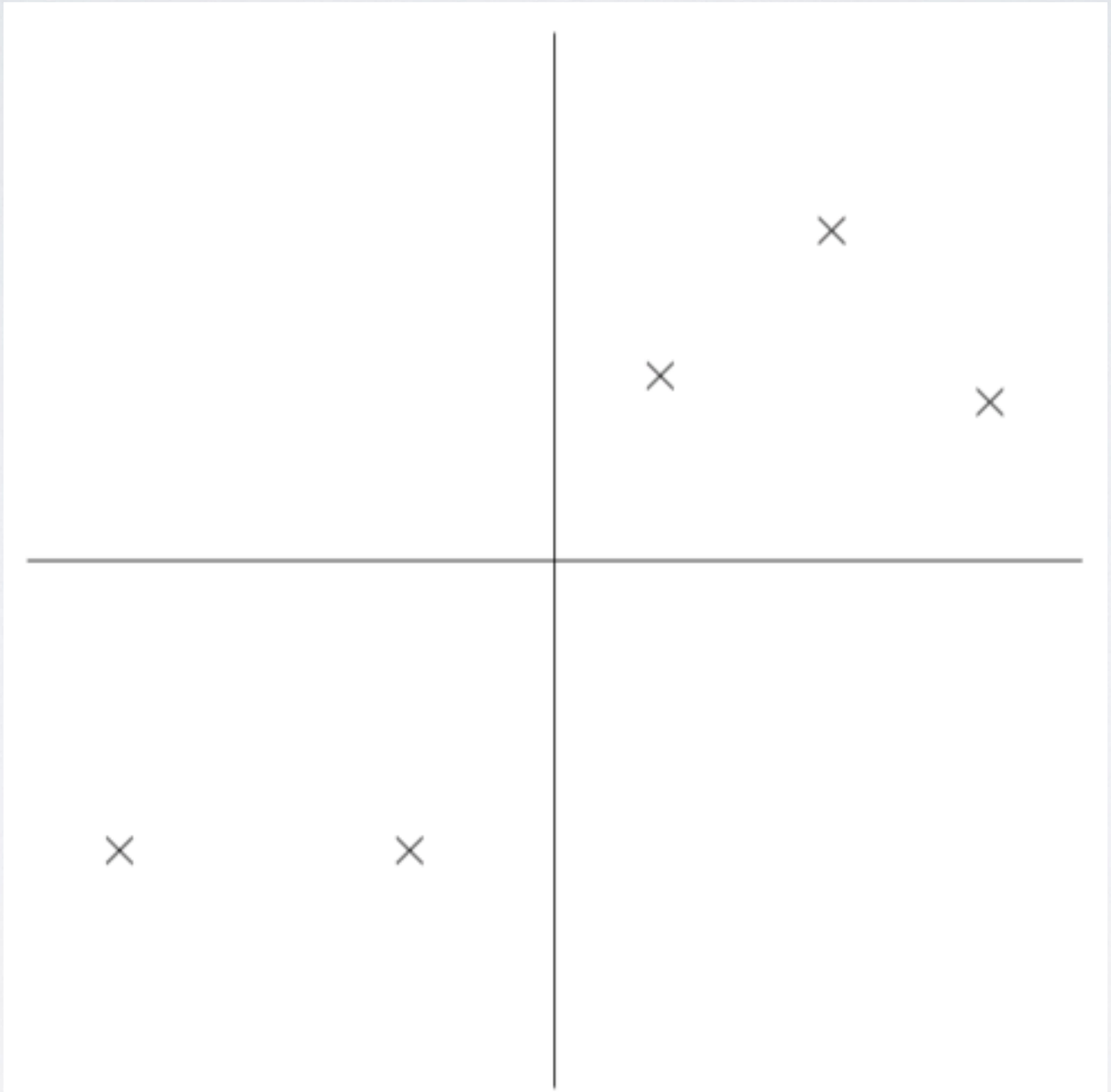
Zero out the mean of the data

Let $\sigma_j^2 = \frac{1}{m}\sum_{i} \left(x_j^{(i)}\right)^2$

Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$

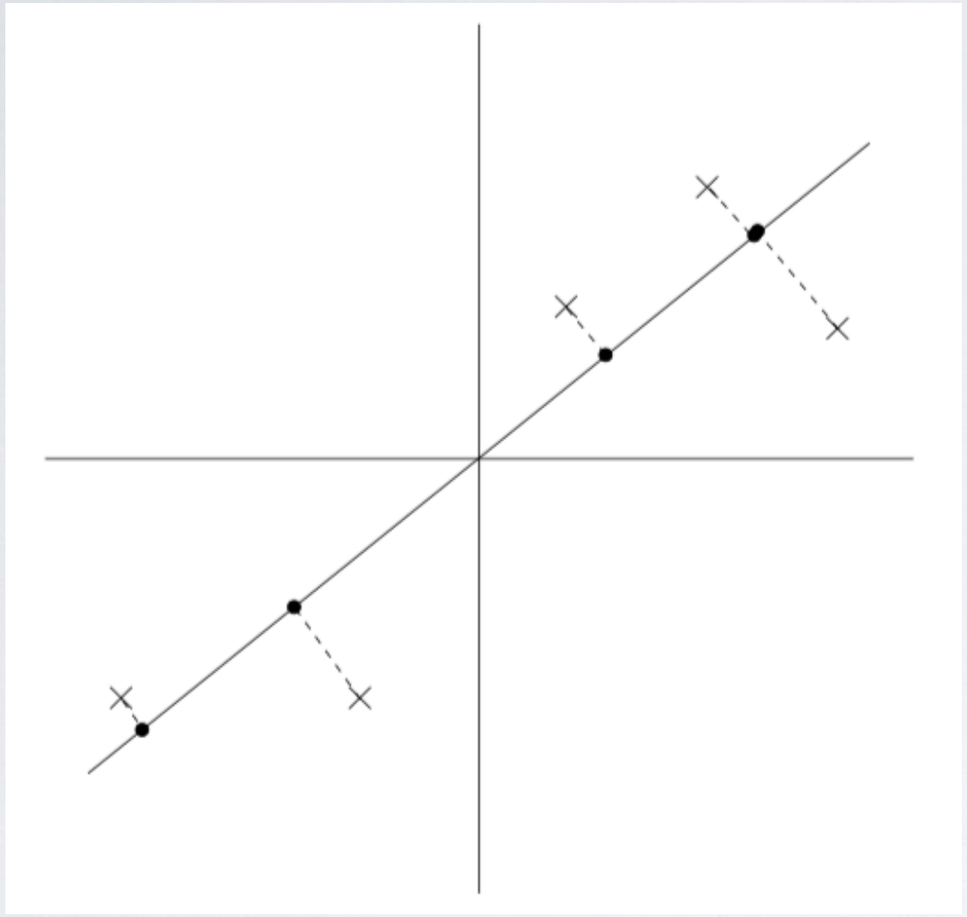
Rescales each coordinate to have unit variance, which ensures that different attributes are all treated on the same scale and makes them comparable

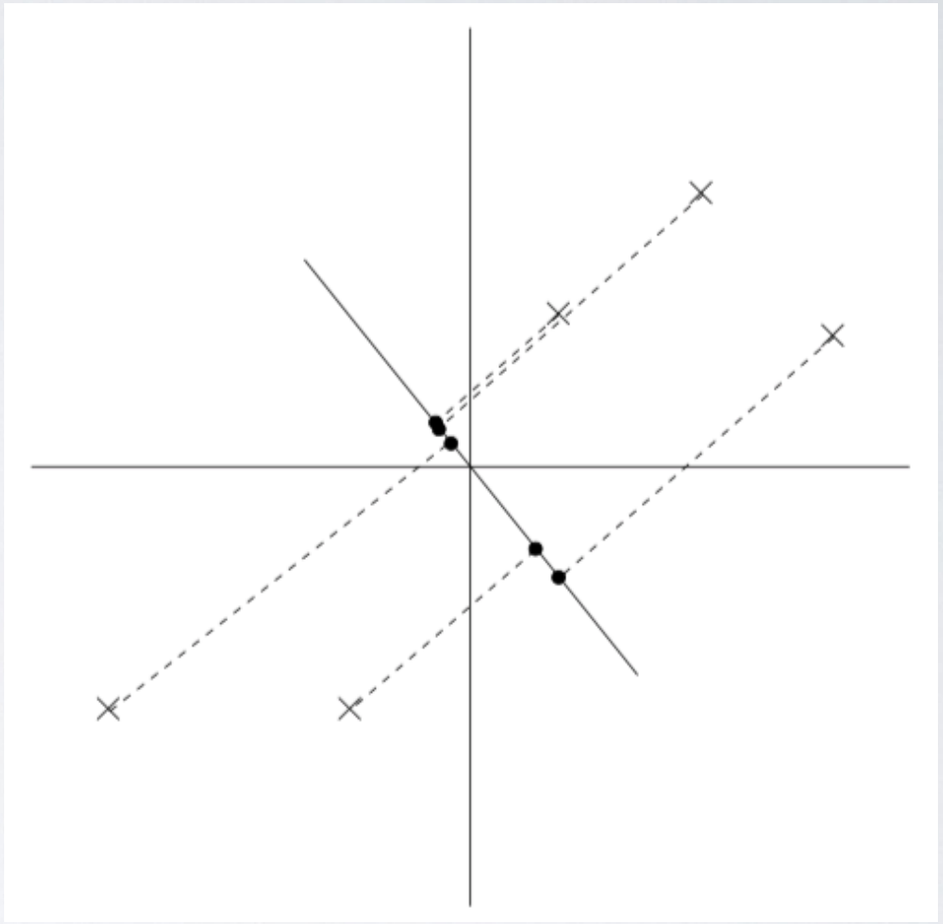
INTUITION I



Which direction retains the most of the data variation?

INTUITION II

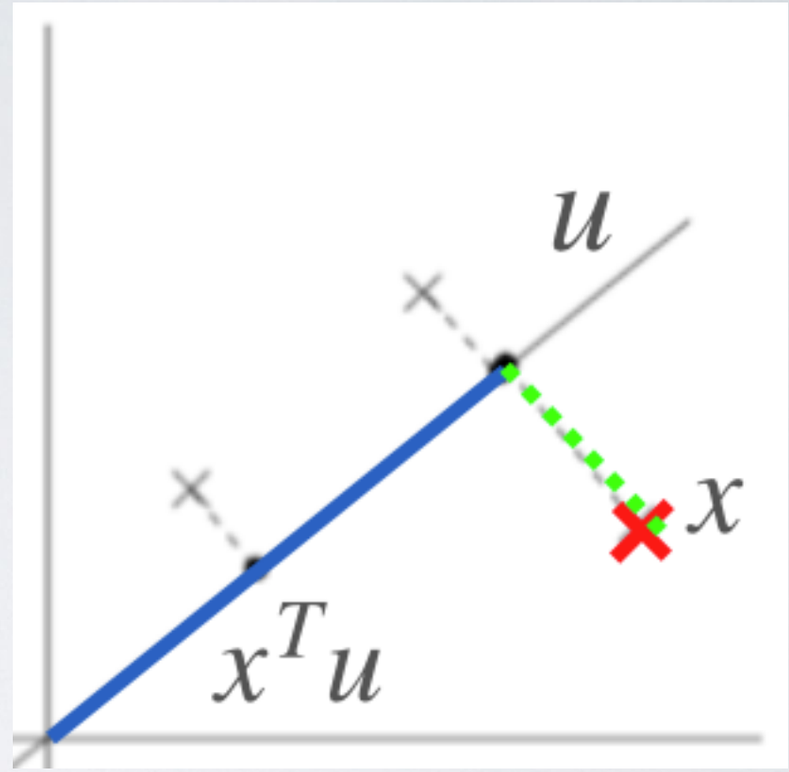




LEFT: Projections have fairly large variance, and are far from zero

RIGHT: Projections have a significantly smaller variance, and are much closer to the origin

- The length of the projection of point x onto u is given by $x^T u$



- To maximize the variance of the projections we would like to choose a unit-length u so as to maximize:

$$\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)^T}u\right)^2$$

$$\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)^T}u\right)^2$$

$$\frac{1}{m}\sum_{i=1}^{m}u^T x^{(i)}x^{(i)^T}u$$

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

Empirical Covariance $$\Sigma = \frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^T}$$

FINAL PROBLEM

$$\underset{u}{\text{argmax}} \quad u^T \Sigma u$$

$$\text{subject to} \quad \|u\|_2 = 1.$$

$$\underset{u}{\text{argmax}} \quad u^T \Sigma u - \lambda(u^T u - 1)$$

Constructing Lagrangian

Setting Derivative to 0

$$\frac{\partial\left(u^T \Sigma u - \lambda(u^T u - 1)\right)}{\partial u} = 2\Sigma u - 2\lambda u = 0$$

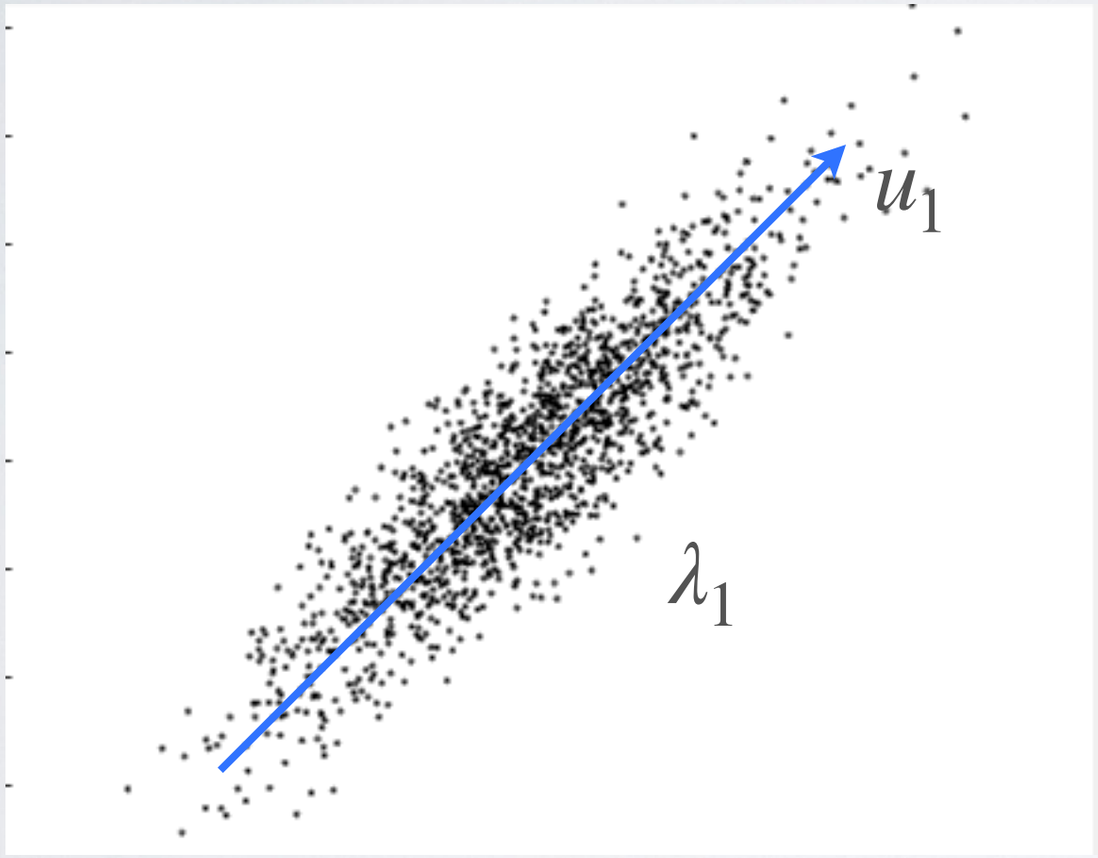
$$\Sigma u = \lambda u \longrightarrow u^T \Sigma u = u^T \lambda u \longrightarrow \underset{||}{u}^T \Sigma u = \lambda u^T u \longrightarrow u^T \Sigma u = \lambda$$

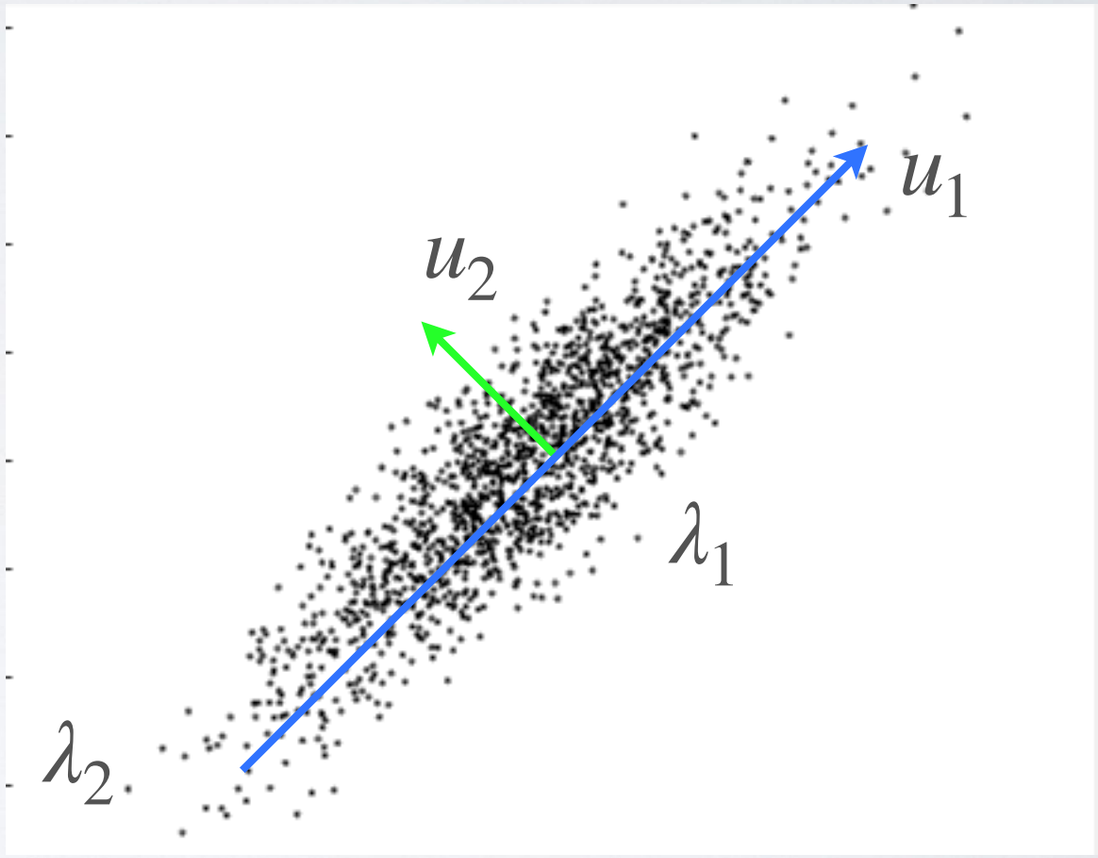
FINAL PROBLEM

$\Sigma u = \lambda u$ ← Solution is the eigenvector of the covariance

$u^T \Sigma u = \lambda$ ← Largest variance is achieved by the eigenvector corresponding to the largest eigenvalue λ

MORE PRINCIPAL COMPONENTS > 1





M PRINCIPAL COMPONENTS

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. If we consider the general case of an M-dimensional projection space, the optimal linear projection for which the variance of the projected data is maximized is now defined by the M eigenvectors of the data covariance matrix S corresponding to the M largest eigenvalues

REPRESENTING THE DATA USING PCA BASIS

$$y^{(i)} = \begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \vdots \\ u_k^T x^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

Thus we see that projection of our data point has $k \ll m$ dimensions

SOME PROPERTIES OF PCA BASIS

PCA basis is orthogonal due to the symmetric nature of the data covariance matrix

PCA & FACES

← Dataset

SOME OF THE EIGEN FACES

THANKS!