EECS 16B   Module 3, Lecture 4   Apr 13, 2021.

Today:

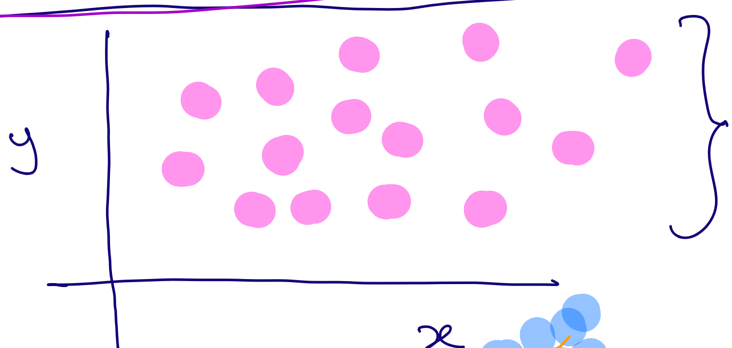
- PCA continued. → Unsupervised classification / Compression.
- Classification (other ways)
  ↳ 16A: Max correlation.
- Linearization

$$Q = \begin{bmatrix} \vec{\ell}_1^T \\ \vec{\ell}_2^T \\ \vdots \\ \vec{\ell}_m^T \end{bmatrix}$$

- $$Q = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T = U \Sigma V^T \qquad Q = \begin{bmatrix} \vec{q}_1 & \cdots & \vec{q}_n \end{bmatrix}$$

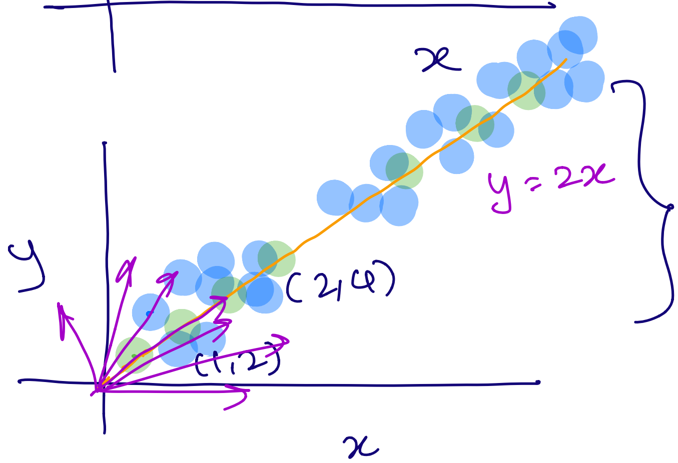
$\vec{u}_i$ are principal components along columns

$\vec{v}_i^T$ —— " —— rows



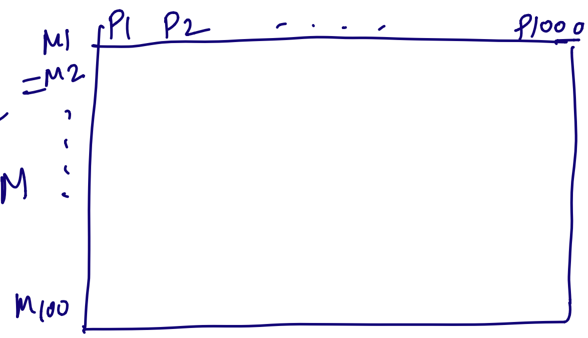
2D

$\mathbb{R}^2$



Underlyin 1D Structure.

$Q =$



Say we care about data along the columns of Q.

To find the lower dimensional structure, find $\vec{w}$, such that, you project $\vec{q_i}$'s onto $\vec{w}$ you have the minimum error. $\|\vec{w}\|^2 = 1$

Different from $\left(\min_{\vec{w}}\right)$

$$\underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{n} \|\vec{q_i} - \langle \vec{q_i}, \vec{w} \rangle \vec{w}\|^2 = \vec{u_1}$$

Similarly? 2nd most imp. component? : $\vec{u_2}$

If instead we did the same derivation, using rows of Q,

$$\underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{m} \|\vec{l_i} - \langle \vec{l_i}, \vec{w} \rangle, \vec{w}\|^2 = \vec{v_1}$$

To find the lower dim. structure of our data, project $\vec{q}_1, \vec{q}_2 \dots \vec{q}_n$ along $\vec{u}_1$

<u>Consider:</u> $\langle \vec{q}_1, \vec{u}_1 \rangle \vec{u}_1, \ \langle \vec{q}_2, \vec{u}_1 \rangle \vec{u}_1 \ \dots$

↳ representation of data along the $\vec{u}_1$ axis.

<u>Summary:</u> $\vec{x}_1, \vec{x}_2, \dots \vec{x}_n$

Find the principal components of this data.

① Arrange data into a matrix

$$X = \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{bmatrix}$$

⟶ "Mean removal" → depending on application. (Out of scope 16B)

② Compute $X = U \Sigma V^T = \sum \sigma_i \vec{u}_i \vec{v}_i^T$

③ For k principal components along the columns : choose $\vec{u}_1, \vec{u}_2 \dots \vec{u}_k$

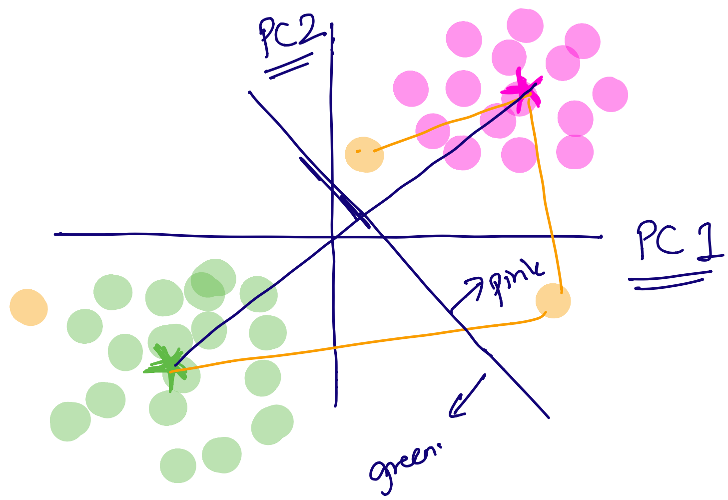
" rows : choose: $\vec{v}_1, \vec{v}_2 \dots \vec{v}_k$

④ (Optional step) : Project data onto $\vec{u}_1, \vec{u}_2, \dots \vec{u}_k$ to get lower dim structure.

→ use this for clustering| classification

Classification.

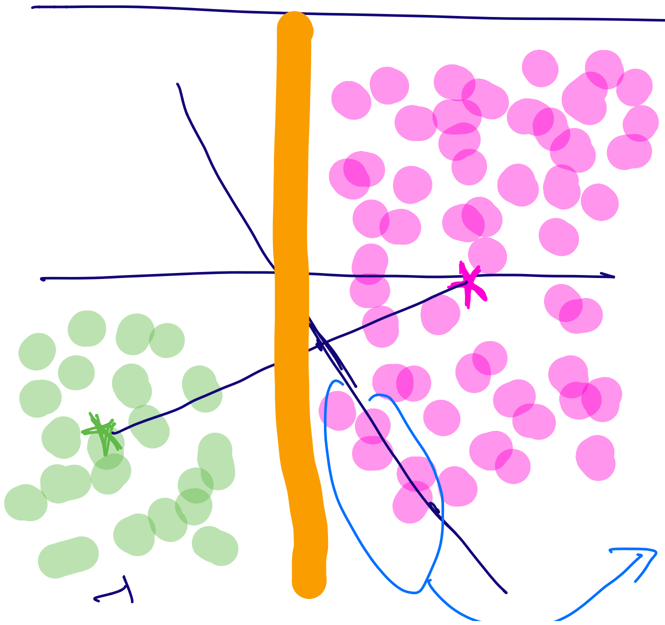
I project data onto top 2 PC. 200 dim data



"find centroids" of data.

Classification using the mean

→ Map new data point to whichever centroid is closer.



→ getting misclassified

In general: $\vec{x}_1, \vec{x}_2 \dots \vec{x}_n$

Want: labels: $l_1, l_2, \dots l_n \in \{+1, -1\}$

Data: $(\vec{x}_1, l_1)$

$(\vec{x}_2, l_2)$

$\vdots$

$(\vec{x}_n, \vec{l}_n)$

Find a classifier:

$$\vec{x}_i^T \cdot \vec{w} + b = l_i$$

"Least squares"

$$\vec{x}_i^T \quad \begin{bmatrix} \vec{x}_1^T & 1 \\ \vec{x}_2^T & 1 \\ \vdots & \\ \vec{x}_n^T & 1 \end{bmatrix} \underbrace{\begin{bmatrix} \vec{w} \\ b \end{bmatrix}}_{\vec{w}} = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}$$

unknowns

$\rightarrow$ Learn $\vec{w}, \vec{b}$.

new point: $\vec{x}_{n+1}^T \vec{w} + b =$

$$\operatorname{sgn}\left[\vec{x}_{n+1} \vec{w} + \vec{b}\right] = l_{n+1}$$

$\text{sgn}(x) = +1 \quad \text{if} \quad x \geq 0$

$\text{sgn}(x) = -1 \quad \text{if} \quad x < 0$

Least squares: minimize: $\displaystyle \min_{\vec{w}} \underbrace{\sum_{i=1}^{n} \|\vec{x}_i^T \vec{w} - l_i\|^2}$

$\vec{x}_i$ "augmented data point"

↳ 1 is appended to each data point.

Classifier $\displaystyle \vec{w}_* = \underbrace{\arg\min_{\vec{w}} \sum_{i=1}^{n} \|\vec{x}_i \vec{w} - l_i\|^2}_{\text{find the best } \vec{w}}$