ANNOUNCEMENTS

1. [5780] Quiz-1 out on canvas, Quiz-2 paper out on website
   (The quiz will be out EOD)
   OPTIONAL!!!

2. Office-hours – schedule on Queue-Me-In, see course website!

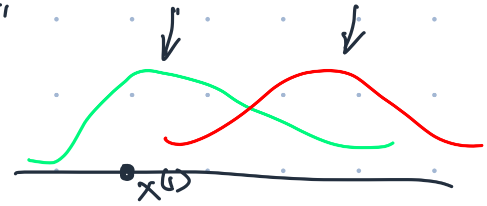
---

while you wait: talk to your 1-nearest-neighbor (Manhattan dist) and ask if they've taken "Intro to Wines" at Cornell!
the Hotelie one

TURN YOUR NON-NOTE-TAKING DEVICES OFF ← we'll give you a device break!

RECAP — mix of Gaussians

ASSUME — similar points form "groups"



REPEAT

1. Randomly choose Gaussians
2. Compute "soft" assignments, prob of $\bullet^{(i)}$ having generated from
3. Recompute the Gaussians

TODAY — PCA

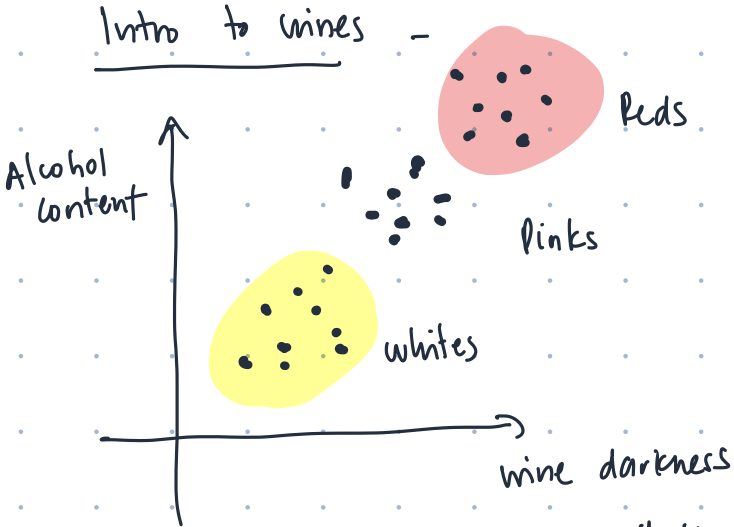
Assume: Data lives in some low-dimensional subspace

CAR data — max speed, turn radius, acceleration, ... — 100 total features

[10] — max speed in mph
[26] — max speed in kph

} data lives in 99 dimensional subspace?

Intro to wines —



Q. a single direction to use to represent this data, which to choose?

≡ Can I find some direction s.t. I can reproduce my data fithfully.

Ritiya says "yes!" →

Matt says "yes" ↑

Max says "↗"

"PCA"

PREPROCESS THE DATA

j – data index

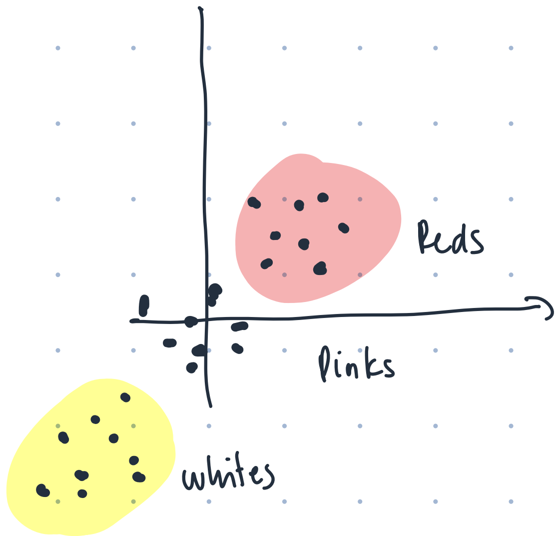
$x^{(j)}$ – j^{th} example,

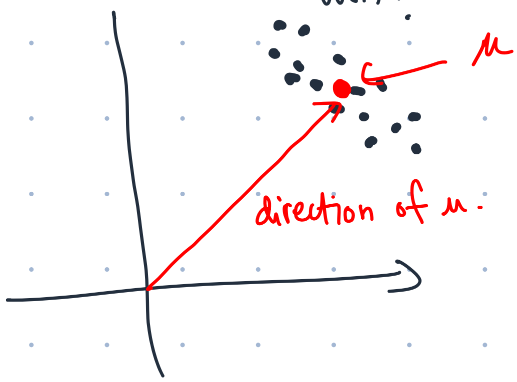
$x_i^{(j)} \leftarrow$ feature 'i'

1. Zero-out the mean

$$\mu \leftarrow \frac{1}{n} \sum_{j=1}^{n} x^{(j)}$$

WHY?





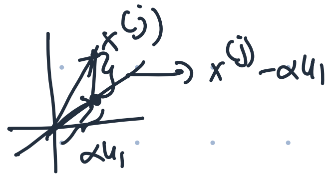
Alc content — w/v % — 60% / 70% ...

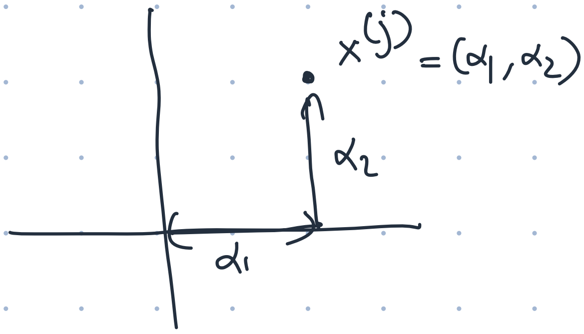
Wine darkness — nm — 590 nm

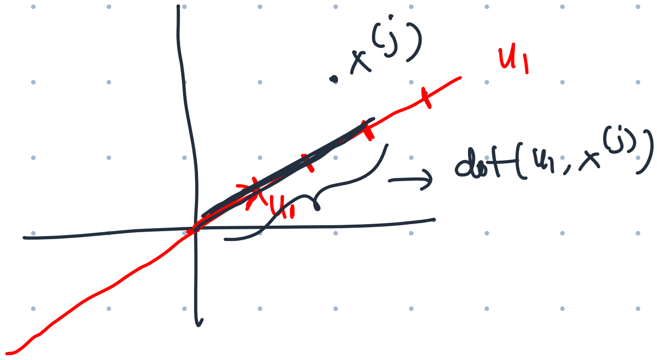
2. (opt) unit variance

$$\sigma_i^2 = \frac{1}{n} \sum_{j=1}^{n} (x^{(j)} - \mu)^2 \quad \Bigg| \quad x_i^{(j)} \leftarrow \frac{x^{(j)}}{\sigma_i}$$

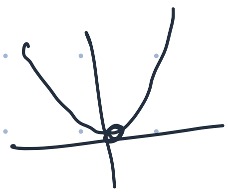
Distances !!! — yay!!







Q: what is the closest point on / to $\cdot\, x^{(j)}$



$$\arg\min_{\alpha} \|x^{(j)} - \alpha u_1\|^2$$

Believe me! – $\|x\|^2 = x^T x$

$$f(\alpha) = \|x^{(j)} - \alpha u_1\|^2$$

$$= (x^{(j)} - \alpha u_1)^T (x^{(j)} - \alpha u_1)$$

$$\nabla_\alpha f = \nabla \left[\underbrace{x^{(j)T} x^{(j)}}_{\to 0} - \alpha u_1^T x^{(j)} - \alpha x^{(j)T} u_1 + \alpha^2 u_1^T u_1 \right]$$

$$= - u_1^T x^{(j)} - x^{(j)T} u_1 + 2\alpha \underbrace{\|u_1\|^2}_{\to 1}$$

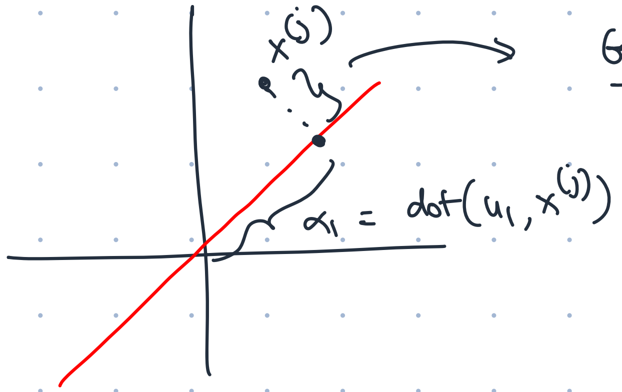
$$= - u_1^T x^{(j)} - x^{(j)T} u_1 + 2\alpha$$

$$= -2 u_1^T x^{(j)} + 2\alpha \overset{\text{set}}{=} 0$$

$$\alpha = u_1^T x^{(j)} \leftarrow \mathrm{dot}(x^{(j)}, u_1)$$

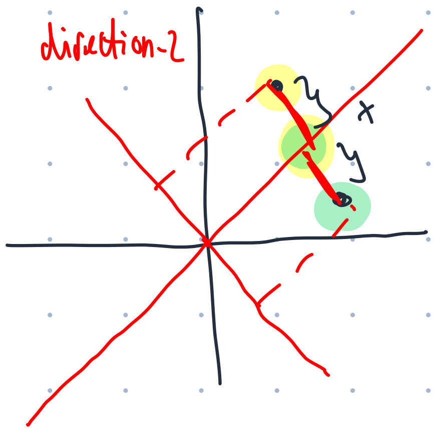
$$u_1^T x^{(j)} = x^{(j)T} u_1$$

$$u_1^T x^{(j)} \triangleq \mathrm{dot}(u_1, x^{(j)})$$



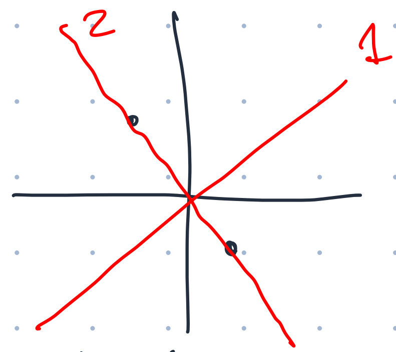
GOAL : MINIMIZE!

Residual $r^{(j)} = \|x^{(j)} - \alpha_1 u_1\|$



direction-1

Griffin says α_1
residual =



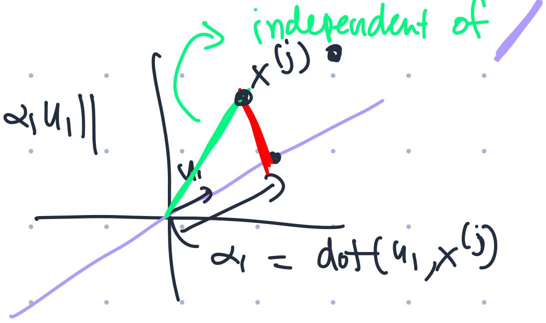
Aditya says "2"
residual - 0

Center the data and do it again,
Which of non-centered/centered is better?

Compute the residuals (geom) and tell me which direction to choose?

PCA —

unexplained $r^{(j)} = \|x^{(j)} - \alpha_1 u_1\|$



$$\arg\min_{u_1} \frac{1}{n}\sum_{j=1}^{n} \|x^{(j)} - \alpha_1 u_1\|^2$$

OR

$$\arg\max_{u_1} \frac{1}{n}\sum_{j=1}^{n} \alpha_1^{(j)\,2} \longrightarrow \text{maxizize "spread"/variance}$$

$\alpha_1 = \text{dot}(u_1, x^{(j)})$

max

$$f(u_1) = \frac{1}{n}\sum_{j=1}^{n} \underline{\underline{(u_1^T x^{(j)})^2}}$$

$\|u_1\| = 1$

subject to

$\|u_1\| = 1$

$\nabla_{u_1} f \overset{\text{set}}{=} 0$

optimize under const — $\|u\| = 1$

$$(u_1^T x^{(j)})^2 = (u_1^T x^{(j)})(u_1^T x^{(j)})$$

$$= (u_1^T x^{(j)})(x^{(j)T} u_1)$$

$u_1^T x^{(j)} = x^{(j)T} u_1$

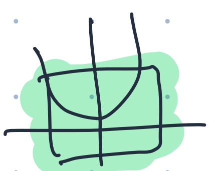
$$= u_1^T x^{(j)} x^{(j)T} u_1$$

$$\frac{1}{n}\sum_{j=1}^{n} u_1^T x^{(j)} x^{(j)T} u_1 = u_1^T \underbrace{\left(\frac{1}{n}\sum_{j=1}^{n} x^{(j)} x^{(j)T}\right)}_{\text{zero-mean data}} u_1$$

empirical covariance matrix

"Σ"

$$= u_1^T \Sigma u_1, \quad \Sigma = \frac{1}{n}\sum_{j=1}^{n} x^{(j)} x^{(j)T}$$



$$\arg\max_{u_1 \in \mathbb{R}^d} u_1^T \Sigma u_1, \text{ subject to } \|u_1\|^2 = 1$$

$\lambda \to -\infty \Rightarrow \max$

Lagrangian —

$$\underset{u_1 \in \mathbb{R}^d}{\arg\max}\ \underset{\lambda}{\arg\min}\ u_1^T \Sigma u_1 - \lambda\left(\|u_1\|^2 - 1\right) = \mathcal{L}(u_1, \lambda)$$

$$\|u_1\|^2 = u_1^T u_1$$

$$\nabla_{u_1} \mathcal{L} \overset{\text{set}}{=} 0$$

$$\nabla_{u_1} = \Sigma u_1 - \lambda u_1 \overset{\text{set}}{=} 0$$

$$\Sigma u_1 = \lambda u_1$$

$$Av = \lambda v$$

This is the e.v. equation!!!

the direction u_1 — min residuals, max α's, OR
the choice I have to make is
the E-vector of Σ

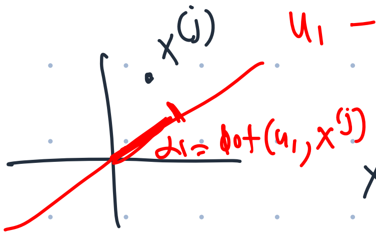
u_1 is the e-vector corresponding to the largest e.v.

PCA

1. center the data
2. compute covariance matrix $\frac{1}{n} \sum x^{(j)} x^{(j)T}$
3. Compute eigen vectors (np.linalg.eig)
4. DONE!!! $\longrightarrow$ gives me "$\underline{u_1}$" — the most optime choice of a direction

$$x^{(j)} \in \mathbb{R}^d \longrightarrow x^{(j)}_{pca} \in \mathbb{R}^k$$

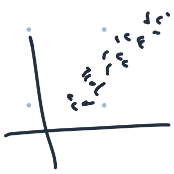
$$\mathbb{R}^2 \qquad\qquad \mathbb{R}^1$$



$$x^{(j)}_{pca} = \alpha_1 = dot(u_1, x^{(j)})$$

$$\frac{1}{n} \sum_{j=1}^{n} x^{(j)} x^{(j)T} = \Sigma \quad \text{"}\underline{\text{cov}}\text{" matrix}$$

PCA de correlates dimensions



$x_1 \uparrow\ x_2 \uparrow$
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma(x^1, x^2) \\ \sigma(x^1, x^2) & \sigma_2^2 \end{bmatrix} \xrightarrow{PCA} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$x_1 \uparrow\ x_2 \downarrow$

RECONSTRUCTION

$$x^{(j)} \longrightarrow x^{(j)} - \mu \xrightarrow{\text{PCA}} x^{(j)}_{pca} \xrightarrow{?} \hat{x}^{(j)}$$

$$x^{(j)}_{pca} = \alpha_1 = \text{dot}(u_1, x^{(j)} - \mu) \equiv \alpha_1 = u_1^T(x^{(j)} - \mu)$$

Reconstruction — multiply OBS by u_1

$$u_1 \alpha_1 = u_1^T x^{(j)} u_1 - u_1^T \mu u_1$$

$$u_1 \alpha_1 = u_1^T x^{(j)} u_1 - \mu$$

$$u_1 \alpha_1 + \mu = \underline{u_1^T x^{(j)} u_1} = \hat{x}^{(j)}$$

Reconstruction