Lecture 10 : ~~PCA, Singular Values,~~ Eigenvalues

①

[[Today we'll (eventually) talk abot "principal component analysis (PCA)", which is another important way of taking high-dim. data points and replacing them by "slightly corrupted" much-lower-dimensional points.

But I don't want to miss out on the algorithmic component, so I'll start talking about that first, even if it seems a little unmotivated.]]

[[First, some linear algebra.]]

[[First first, an analogy.]] Graphs: undirected ☺
directed ☹ [[ugh, annoying to deal with, but sometimes you must]]

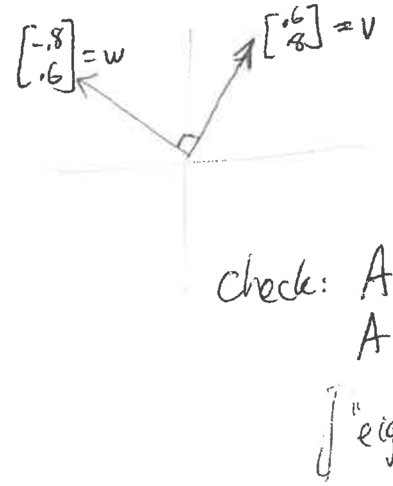
Matrices: symmetric $(A = A^T)$ ☺
nonsymmetric (or rectangular!) ☹ [[again annoying, try not to deal with them if you can]]

Why is a matrix? Operation that transforms vectors

e.g. $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ = "Stretch by factor 3 in 1st coord,
" " " 2 " 2nd coord". $A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x \\ 2y \end{bmatrix}$

[[symmetric ☺]]

$A = \begin{bmatrix} 41 & -12 \\ -12 & 34 \end{bmatrix} = ?$



Stretch by factor 25 along v-axis, factor 50 along w-axis.

Check: $Av = 25v$
$Aw = 50w$

[["eigenvectors/vals"]]

[[How do we know/find this??]]

$A = \begin{bmatrix} .94 & -.34 \\ .34 & .94 \end{bmatrix} = ?$

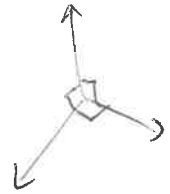
Rotates vecs by ≈ 20° c.c.w.

[[not symmetric]]

[[not so bad, but still a bit annoying]]

Linear Algebra fact: Let A be a $D \times D$ symmetric matrix.

Then their is some orthogonal basis of unit vectors $\vec{v}_1, \ldots, \vec{v}_D$ ~~such that A "~~ and real #'s $\lambda_1, \ldots, \lambda_D$



such that $A \equiv$ "stretch by λ_j in direction $\vec{v}_j$ $\forall j$"

"eigenvectors"

"eigenvalues"

[[negative is possible, means reflecting. ≥ 0 is nicer of course...]]

[[So symm. matrices are really simple: just "stretchers", in some basis]]

Q: How to find $\vec{v}_j$'s and λ_j's?

A1: Type "eigs(A)" in Matlab ;-) [[But how does it do it?]].

A2: [[Some Gaussian-elimination-type alg. you learn in linear algebra class. Time is $\Theta(D^3)$, yuck.]]

A3: Assume, for simplicity, λ_j's all ≥ 0. [["Positive semidefinite matrix." Not easy to tell, given, A, but also not hard to handle. But in future application to PCA, it will be true.]]

(Fuzzy) Alg:

```
① Pick a random vector $\vec{u}_0$, make it unit-length
    by replacing with $\vec{u}_0 / \|\vec{u}_0\|$
② For $i = 1, 2, 3, \ldots$
    $\vec{u}_i := A\vec{u}_{i-1}$, then make it unit
    If $\vec{u}_i \approx \vec{u}_{i-1}$,   // up to some precision
        break.
③ $\vec{v}_{max} := \vec{u}_i$   // found "first" eigenvec.
  // $A\vec{v}_{max} \approx \lambda_{max}\vec{v}_{max}$ for some scalar $\lambda_1$; calculate it.
```

[[people often use Gaussians so it's rotationally symm., but doesn't matter much; ±1 coords ok]]

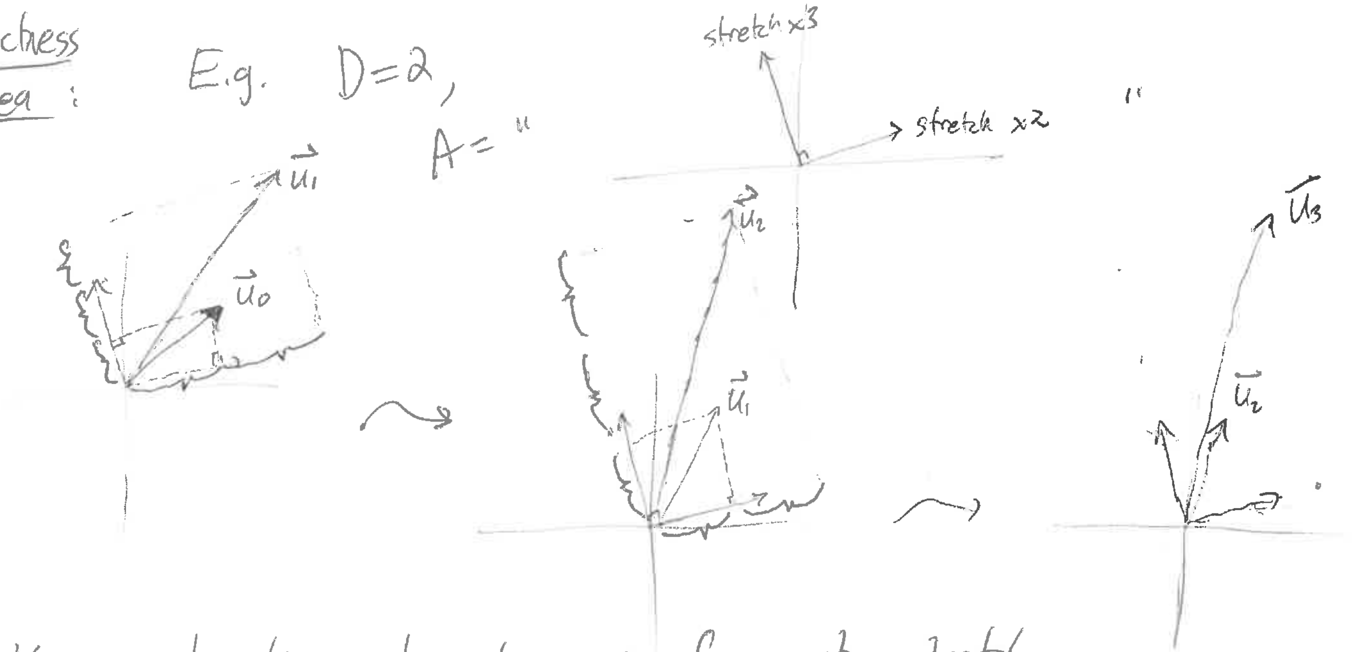
This (heuristically) finds the largest scale factor (eigenval.) λ_{max} and v_1. [["Repeat?"]] [[We'll come back to that.]]

Time? $O(D^2) \cdot (\#\text{ iterations})$ ⟦Usually small, as we'll see.⟧ ③

⟦Kind of a linear-time alg., since input A is a $D \times D$ matrix, D^2 #'s.⟧

Correctness idea: E.g. $D=2$,

$A =$ "



Keeps trending toward axis of most stretch.

Analysis: ⟦Though the alg. doesn't know $\overset{\text{(basis)}}{\text{eigenvectors}}$ $\vec{v}_1, \ldots \vec{v}_D$, we can do the analysis in that basis.
So... may as well assume it's the "usual" axes!⟧

- W.L.O.G. $\vec{v}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j$ for $j = 1 \ldots D$. So $A \begin{bmatrix} \\ x \\ \\ \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \vdots \\ \lambda_D x_D \end{bmatrix}$.

- For analysis, don't need to keep unit-izing...

So $\vec{u}_0 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$, $\vec{u}_1 = \begin{bmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \vdots \\ \lambda_D x_D \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} \lambda_1^2 x_1 \\ \vdots \\ \lambda_D^2 x_D \end{bmatrix}$, ..., $\vec{u}_t = \begin{bmatrix} \lambda_1^t x_1 \\ \vdots \\ \lambda_D^t x_D \end{bmatrix}$

random unit length → heuristically, all x_i's $\pm\frac{1}{\sqrt{D}}$-ish.

Imagine, e.g. $\lambda_3 = 10$, all other λ_i's ≤ 9.

After t iters, coord. #3 of $\vec{u}_t$ is $10^t \cdot \underset{\uparrow \atop \pm O(\frac{1}{\sqrt{D}})}{x_3}$, all other coords $\leq 9^t \cdot O(\frac{1}{\sqrt{D}})$ in abs. val.

e.g. $\vec{u}_t = \begin{bmatrix} 8.4^t \cdot \frac{3}{\sqrt{D}} \\ 9^t \cdot \left(\frac{-.2}{\sqrt{D}}\right) \\ 10^t \cdot \frac{.4}{\sqrt{D}} \\ \vdots \\ \vdots \end{bmatrix}$

Say we unit-ize.

$$\|\vec{u}_t\|^2 = \underbrace{\left(8.4^t \cdot \frac{3}{\sqrt{D}}\right)^2}_{\le 81^t \cdot O\left(\frac{1}{D}\right)} + \underbrace{\left(9^t \cdot \left(\frac{-.2}{\sqrt{D}}\right)\right)^2}_{\le 81^t \cdot \frac{.04}{D}} + \underbrace{\left(10^t \cdot \frac{.4}{\sqrt{D}}\right)^2}_{100^t \cdot \frac{.16}{D}} + \cdots \quad \cdots$$

All coords other than 3^{rd} add to $\le O(81^t)$

3^{rd} coord is $\gg \cancel{O(81^t)}$

$$\Leftrightarrow \quad 100^t \frac{\cdot \text{const}}{D} \gg 81^t$$

$$\Leftrightarrow \quad \left(\frac{10}{9}\right)^{2t} \gg D$$

$$\Leftrightarrow \quad 2t \cdot \log\left(\frac{10}{9}\right) \gg \log D.$$

The ratio $\frac{9}{10}$ called the "spectral gap".

$$\Leftrightarrow \quad t \gg \log D$$ in this case where $\frac{10}{9} \ge 1 +$ "constant".

In this case, after unit-izing $\vec{u}_t$, looks like $\begin{bmatrix} 0.00\ldots \\ 0.00\ldots \\ .999\ldots \\ 0.00\ldots \\ \vdots \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$

We found the direction of greatest stretch! "$\vec{v}_{max}$"

Mult by A once more $\rightarrow$ $\begin{bmatrix} 0 \\ 0 \\ 10 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ $\leftarrow \lambda_{max}$.

time: $O(D^2 \underbrace{\log D}_{\text{iters}})$

//

This worked well because we assumed λ_{max} (10) $> \lambda_{2nd\,max}$ (9) by a decent factor.

What if $\lambda_{max} = (1+\varepsilon)\lambda_{2nd\,max}$?

\# iters needed: $t \gg \dfrac{\log D}{\log\left(\frac{\lambda_{max}}{\lambda_{2nd}}\right)} = \dfrac{\log D}{\log(1+\varepsilon)}$.

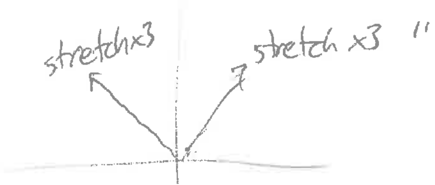
How much is $\log(1+\varepsilon)$? $1+\varepsilon \approx e^{\varepsilon}$

$\therefore \ln(1+\varepsilon) \approx \varepsilon$

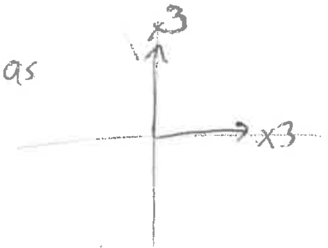
$\Rightarrow$ Need $t \gg \dfrac{\log D}{\varepsilon}$. [[It's even possible $\varepsilon = 0$...]]

[[But actually, it's sorta not that bad.]]

Imagine $\lambda_{max} = \lambda_{2nd}$. E.g. $A =$ "



This is ~~same~~ same as

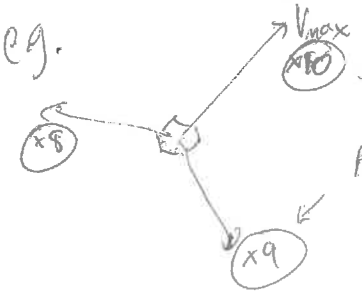


or any basis. So "correct" axes not even defined.

Alg. will just return some vector in the subspace spanned by all those axes with λ's tied for λ_{max}.

Repeat? [[This just finds the direction of max stretch.
Want to find whole basis, all stretch factors
Or maybe, at least the top few largest stretches.]]

eg.



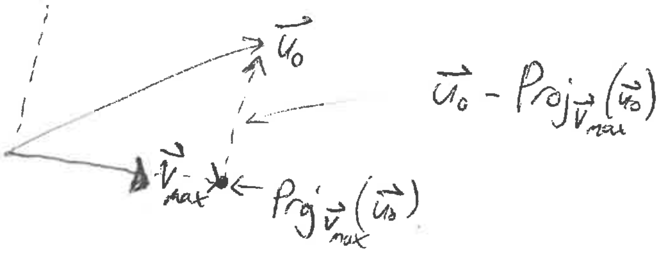
} we just figured this out.

Now want to figure this out.

Idea: Repeat alg., but stay perpendicular to axis (axes) found already stay in subspace of mystery axes.

→ If you start with $\vec{u}_0$ perpendicular to $\vec{V}_{max}$ ["0" in the direction $\vec{V}_{max}$], all $\vec{u}_t$'s will be too, and you'll find $\vec{V}_{2nd\,max}$, $\lambda_{2nd\,max}$.

[How to achieve that?]



Init. · Pick $\vec{u}_0$ at random.

· Replace with

$$\vec{u}_0 - \underbrace{Proj_{\vec{v}_{max}}(\vec{u}_0)}_{(\underbrace{\text{len. of } \vec{u}_0 \text{ projected onto } \vec{v}_{max}}_{\vec{u}_0 \bullet \vec{V}_{max}}) \cdot \vec{V}_{max}}$$

[$\vec{V}_{max}$ a unit vector]

$$\left(= \|\vec{u}_0\| \cdot \underbrace{\|\vec{v}_{max}\|}_{1} \cdot \cos\theta \right)$$

[Can repeat, projecting away all the eigenvectors you've found so far. Finds "top k eigenvectors" in time $O(k D^2 \log D)$]

[Why? When do you ever have this situation??]

Let $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^D$ be "data items

Let

$$X = n\updownarrow \underbrace{\begin{bmatrix} —x^{(1)}— \\ —x^{(2)}— \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}}_{\leftarrow D \rightarrow}$$

[Like last time]. Let $A = X^T X = \begin{bmatrix} & \\ & \end{bmatrix}$

$= D \times D$ matrix

$$A^T = (X^T X)^T = X^T X^{TT} = A$$

$\Rightarrow A$ symmetric & square ☺

Also: all λ's ≥ 0 [easy to show]

V's & λ's for A give crucial info about $x^{(i)}$'s.....