16B Prof Anant Sahai

Announcements: (1) Beta version of grade status out
(2) TI Design Contest
(3) Another Extra-Credit Opp coming....

Today: Finish SVD
Dimensionality Reduction & PCA
Using this for classification

3 Forms of SVD of $m\{\overset{n}{\overbrace{A}} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$

1. Full SVD $A = m\{\overset{m}{\overbrace{U}}\, m\{\overset{n}{\overbrace{\Sigma}}\, \underbrace{V^T}_{n}\}n$ where $U = [\overset{r}{\overbrace{U_1}}\ \overset{m-r}{\overbrace{U_2}}]$

"Polar Coordinate View" of the matrix A

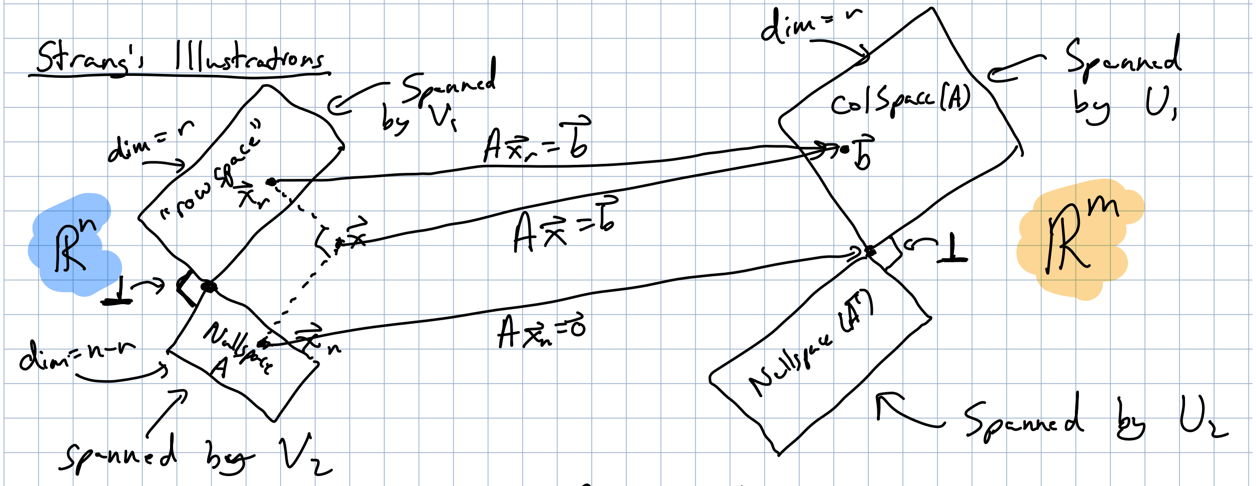
U, V both orthonormal bases. $V^T = \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \begin{matrix} \}r \\ \}n-r \end{matrix}$ Basis for Nullspace(A)

e.g. $\Sigma = \left[\begin{matrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r & & \\ & & & & 0 & \\ & & & & & \ddots \end{matrix} \quad 0 \right]$ or $\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \\ & & & \sigma_r & \\ & & & & 0 \\ & & & & & \ddots & 0 \\ & & 0 & & \end{bmatrix}$

2. $\boxed{\text{Outer-product form } A = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^T}$

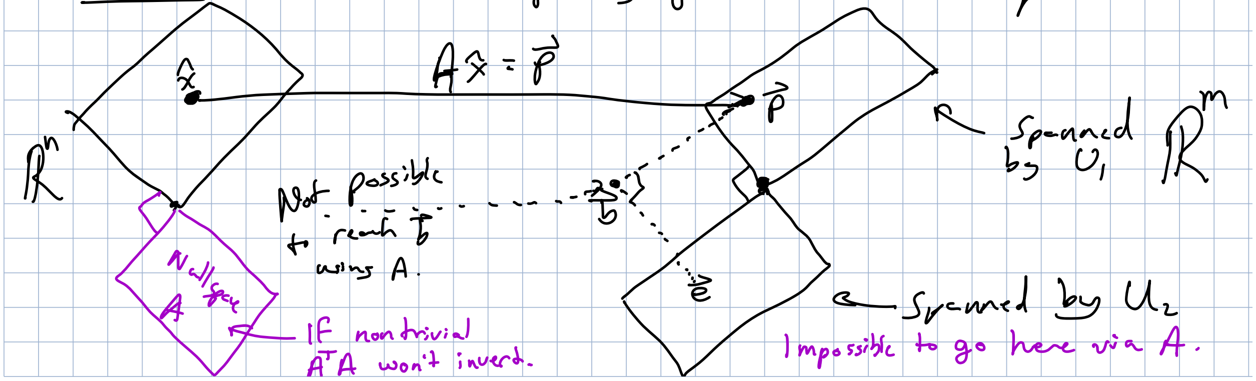
3. Compact form $A = \overset{r}{\overbrace{U_1}}\, r\{\overset{r}{\overbrace{\Sigma_c}}\, V_1^T\}r$

Strang's Illustrations

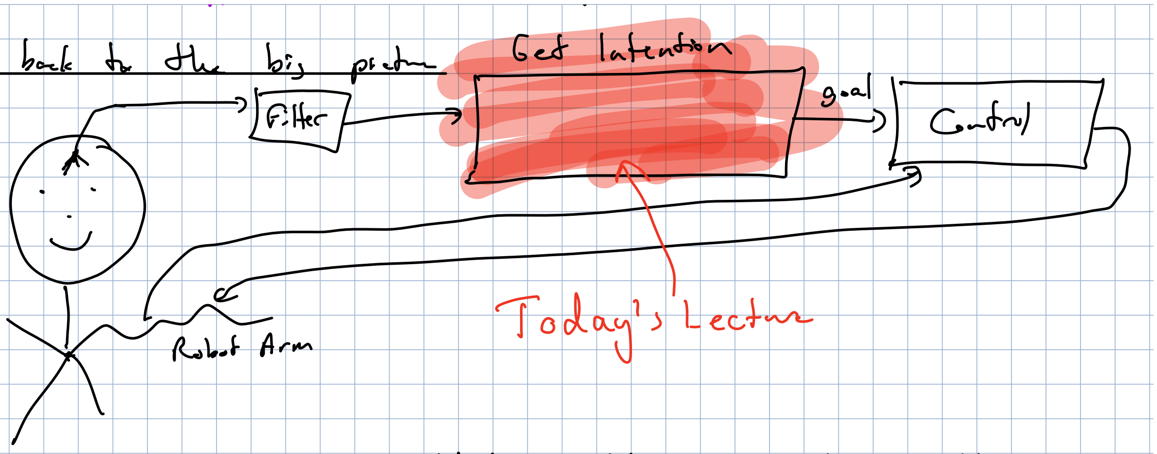


Min-Norm Problem Illustration.

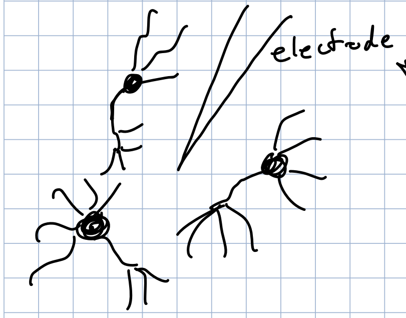
Exercise: Draw the corresponding picture for least-squares.



Come back to the big picture



For simplicity: Suppose that in the person's brain, there are different neurons that correspond to different goals



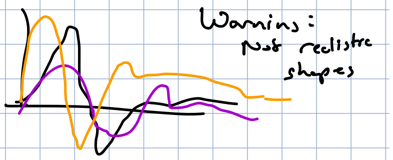
Picking up signals from multiple neurons.

Picks up a waveform

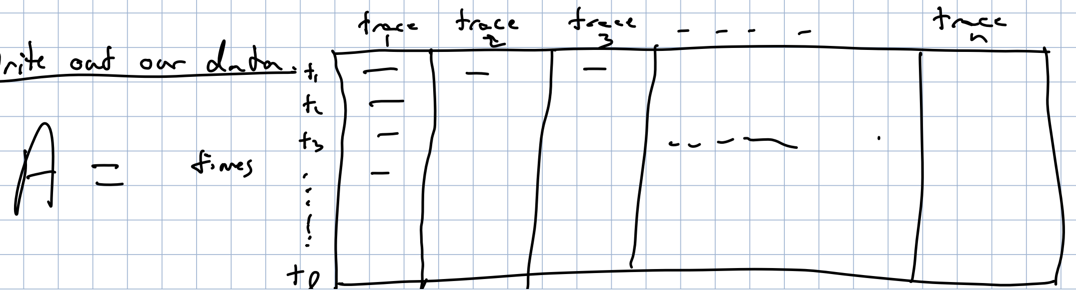
Need to pull apart the different neurons signals.

We don't know
- # of total neurons being picked up. (How many are in range)
- What exactly the firing pattern/spike signal from each of these neurons looks like.

⟹ Assume that the spike signals from diff neurons look different as waveforms.


Warning: Not realistic shapes

Write out our data.



Let $\vec{r}_j$ be the j^{th} recorded trace.

Have a block of recorded data.

If we assume that there are actually k neurons in range and each has a specific signature waveform $\vec{s}_i$.

Then $\vec{r}_j = \sum_{i=1}^{k} w_{i,j} \cdot \vec{s}_i$

↖Some weights. Saying $\vec{r}_j$ is a linear combo of $\{\vec{s}_i\}$

$$\ell\left\{\overbrace{A}^{n} = \ell\left\{\underset{\overset{\shortparallel}{S}}{\overbrace{\boxed{\vec{s}_1\,\vec{s}_2 \cdots \vec{s}_k}}^{k}}\ \overbrace{\boxed{\quad W \quad}}^{n}\right\}k\right. \right.$$

⟵ A has rank k.

$k \ll n$

$k \ll \ell$

If I took the SVD of A, $A = \sum_{i=1}^{k} \sigma_i \vec{u}_i \vec{v}_i^T$

In particular, I'd know k.

Reality Check: In many cases, the singular values look like

$\sigma_1 = 100 \quad \sigma_2 = 90 \quad \sigma_3 = 50 \quad \sigma_4 = 2 \quad \sigma_5 = 1.1 \quad \sigma_6 = 0.3$

$\ldots \ \sigma_{12} = 0.09$

You won't get zeros.

But you will get small numbers.

$\sigma_{18} = 0.001$

$\sigma_{19} = 0.0001$

$\sigma_{20} \ldots$

$\sigma_{30} = 10^{-8}$

$$A = 100 \cdot \vec{u}_1 \vec{v}_1^T + 90 \cdot \vec{u}_2 \vec{v}_2^T + 50\, \vec{u}_3 \vec{v}_3^T + 2 \cdot - \ldots$$

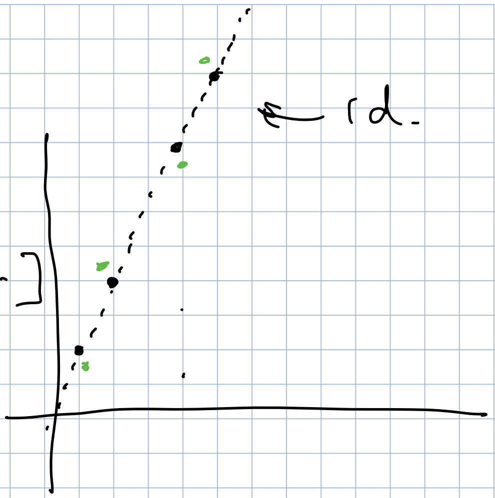
Take the perspective of Taylor Series (Calculus) .. Bode Plots

Let's approximate this. Just keep the big terms.

Quick 2d example.

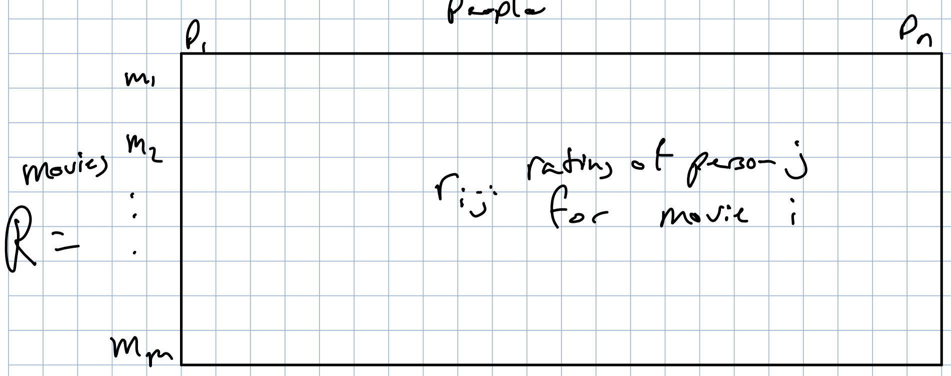
$$A = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 8 & 10 \end{bmatrix}$$

$$SVD(A) = \sqrt{230} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{46}} & \sqrt{\frac{4}{46}} & \cdots \end{bmatrix}$$



Appreciate the audacity of what we're doing.
Our A is just a block of data.

Illustrative conceptual example: Movie Ratings.



$r_{ij} \in \mathbb{R}$
real number score.

16- philosophy: Make a model,
Learn the model from data,
Use the model to make predictions.

Model for movies: Movie i has 4 # associated to it.

e.g. a_i : Action Score

b_i : Bechdel Score

c_i : Comedy Score

d_i : Drama Score

Model for People: Person j has 4 sensitivities

α_j : sensitivity to action

β_j : " " Bechdel score

γ_j : " " Comedy

δ_j : " " Drama.

Model for Ratings:

$$r_{ij} = \alpha_j a_i + \beta_j b_i + \gamma_j \cdot c_i + \delta_j \cdot d_i$$

Challenge: We only see ratings r_{ij}.

Can we learn our model from the data?

$$m\left\{\overbrace{R}^{n}\right. = \vec{a}\,\vec{\alpha}^T + \vec{b}\,\vec{\beta}^T + \vec{c}\,\vec{\gamma}^T + \vec{d}\,\vec{\delta}^T + \underset{\underset{\underset{\text{“hopefully small”}}{\uparrow}}{\text{“disturbance”}}}{\underset{\uparrow}{W}}$$

Where $\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ and similarly for $\vec{b}, \vec{c}, \vec{d}$

$$\vec{\alpha}^T = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n]$$

R has rank 4.

$$R = \underbrace{\left(\sum_{k=1}^{4} \sigma_k \vec{u}_k \vec{v}_k^T\right)}_{\text{Keep these for the model.}} + \underset{\underset{\text{Hope all of these are very small}}{\uparrow}}{\sum_{k=5}^{\min(n,m)} \sigma_k \vec{u}_k \vec{v}_k^T} \quad \text{By SVD.}$$

$$\text{Wanted } R = \underbrace{\vec{a}\,\vec{\alpha}^T + \cdots + \vec{d}\,\vec{\delta}^T}_{4 \text{ terms}}$$

Q: Do we now know our model??

A: <u>No.</u> Fundamentally some ambiguity exists.

1) Could say $\vec{a} = \vec{u}_1$, then $\vec{\alpha}^T = \sigma_1 \vec{v}_1^T$

or $\vec{a} = \sigma_1 \vec{u}_1$, then $\vec{a}^T = \vec{v}_1^T$

2) Similarly: We could have chosen $\vec{a}, \vec{b}$ to be mixtures of $\vec{u}_1$ & $\vec{u}_2$.

All we really know is that we have a 4-D subspace.

Can Modify our <u>model</u> : Each movie has a vector $\vec{g}$ of qualities associated
to reflect what is possible

Each person has a vector $\vec{s}$ of sensitivities.

$$r(\vec{g}, \vec{s}) = \sum_{k=1}^{4} \sigma_k \, g_k \, s_k$$

← numbers

↖ Pull this out instead of trying to allocate it.

Now our learned model has
$$g_i = \begin{bmatrix} (\vec{u}_1)_i \\ (\vec{u}_2)_i \\ (\vec{u}_3)_i \\ (\vec{u}_4)_i \end{bmatrix}$$

Recall $\vec{u}$'s are m long

$\vec{v}$'s are n long.

$$\vec{s}_j = \begin{bmatrix} (\vec{v}_1)_j \\ (\vec{v}_2)_j \\ (\vec{v}_3)_j \\ (\vec{v}_4)_j \end{bmatrix}$$

<u>Next time:</u> Using the model to do prediction?