EECS16A DIS14A

Learning Objectives

- How to set up least squares problems (Identify $A, \vec{x}, \vec{b}$)
- Condition on problem to set up as a least squares problem
  - ↳ What kind of models? (When can we express as $A\vec{x}$?)
- Condition on whether least squares has a unique solution
  - ↳ Also to be covered in lecture tomorrow/picked up on from last week's lecture (Has to do with A, A^TA)
- If time, some ML implications

Music

- Esperanza Spalding
  Black Gold
- Robert Glasper
  Afro Blue

EECS 16A Designing Information Devices and Systems I
Fall 2020 Discussion 14A

1. **Polynomial Fitting**

Let's try an example. Say we know that the output, y, is a quartic polynomial in x. This means that we know that y and x are related as follows:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$$

We're also given the following observations:

	x	y	
$x_0 \rightarrow$	0.0	24.0	$\leftarrow y_0$
$x_1 \rightarrow$	0.5	6.61	$\leftarrow y_1$
$x_2 \rightarrow$	1.0	0.0	$\leftarrow y_2$
⋮	1.5	−0.95	⋮
	2.0	0.07	
	2.5	0.73	
	3.0	−0.12	
	3.5	−0.83	
	4.0	−0.04	
	4.5	6.42	

$\rightarrow b$

(a) What are the unknowns in this question? What are we trying to solve for?

$\Rightarrow \underline{a_0, a_1, a_2, a_3, a_4}$

(b) Can you write an equation corresponding to the first observation (x_0, y_0), in terms of a_0, a_1, a_2, a_3, and a_4? What does this equation look like? Is it linear in the unknowns?

$(x_0, y_0) = (0, 24) \rightarrow y_0 = a_0 + a_1 x_0 + a_2 x_0^2 + a_3 x_0^3 + a_4 x_0^4$

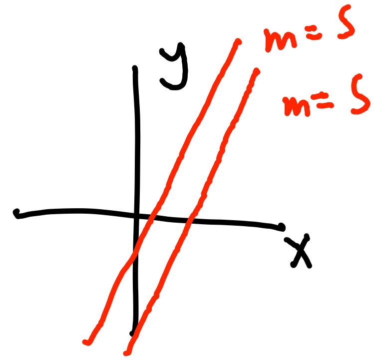
$24 = a_0 + a_1 \cdot 0 + a_2 \cdot 0 + \ldots$

$\rightarrow 24 = a_0$ Linear? Yes - linear comb. of a_i's

(c) Now, write a system of equations in terms of a_0, a_1, a_2, a_3, and a_4 using *all of the observations*.

$$A\vec{x} = \vec{b} \qquad \vec{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_4 \end{bmatrix}$$

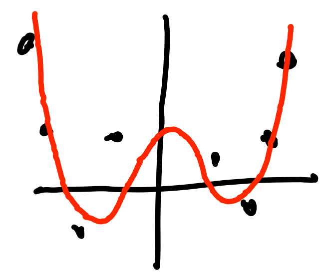
$$\underbrace{\begin{bmatrix} 1 & x_0 & x_0^2 & x_0^3 & x_0^4 \\ 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ & & \vdots & & \\ 1 & x_4 & x_4^2 & x_4^3 & x_4^4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}}_{\vec{x}} \approx \underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_4 \end{bmatrix}}_{\vec{b}}$$



$$y = \underbrace{1 \cdot a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4}_{\text{linear in the } a_i\text{'s}}$$

Substitute values of x's and y's

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ 1 & 1 & 1 & 1 & 1 \\ \vdots & & & & \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ \vdots \\ a_4 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} 2.4 \\ 6.61 \\ 0 \\ \vdots \end{bmatrix}}_{\vec{b}}$$



error: $\|A\vec{x} - \vec{b}\|^2$

$$\hat{x} = \underline{(A^T A)}^{-1} A^T \vec{b}$$

↳ inverse not always defined

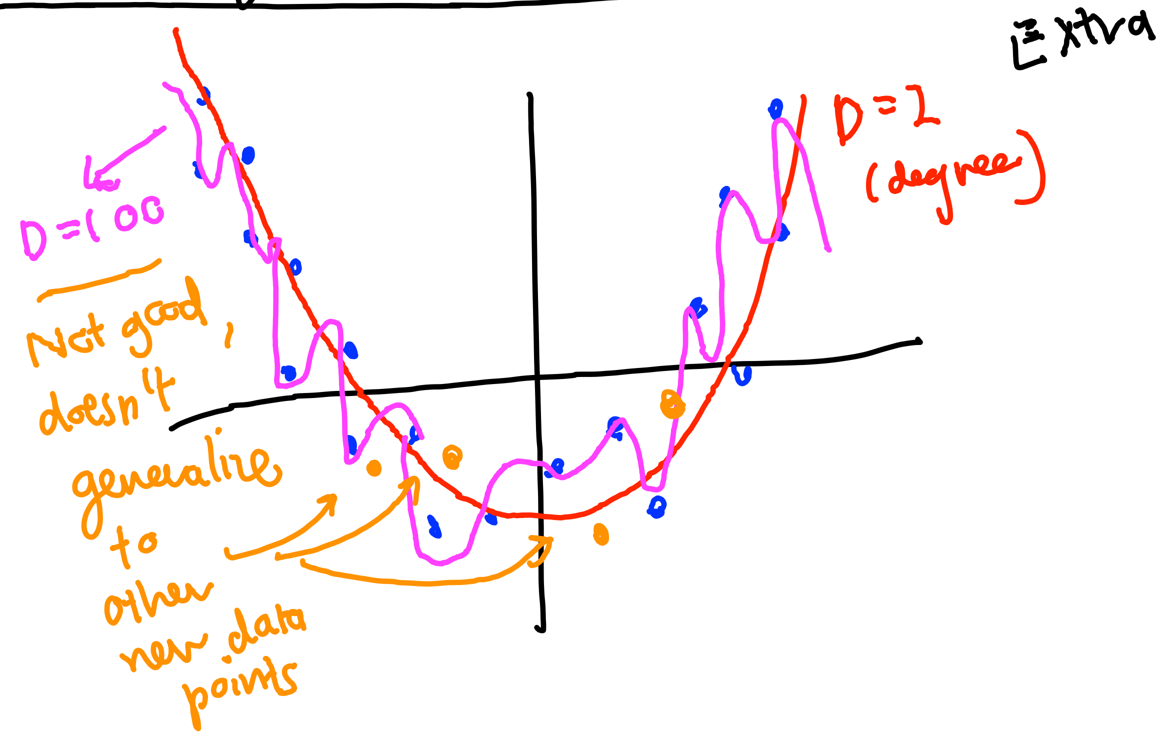
need to know $A^T A$ is invertible

$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ ← no inverse

$A^T A \vec{x} = A^T \vec{b}$ ⇐ always has a solution

(d) Finally, solve for a_0, a_1, a_2, a_3, and a_4 using IPython. You have now found the quartic polynomial that best fits the data!

Look @ Notebook! $a_0 \approx 2.4$! $a_1 = \dots$



2. Building a classifier (Final - Fall 2019)

Least squares are often used in practice to classify data. In this scenario, we would like to develop a classifier to classify points based on their distance from the origin.

You are presented with the following data. Each data point $\vec{d}_i^{\,T} = [x_i\ y_i]^T$ has the corresponding label $l_i \in \{-1, 1\}$.

x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 1: *
Labels for data you are classifying

(a) (6 points) You want to build a model to understand the data. You first consider a linear model, i.e. you want to find $\alpha, \beta, \gamma \in \mathbb{R}$ such that $l_i \approx \alpha x_i + \beta y_i + \gamma$.
Set up a least squares problem to solve for α, β and γ. If this problem is solvable, solve it, i.e. find the best values for α, β, γ. If it is not solvable, justify why.

$$X = \begin{array}{c} \begin{array}{ccc} x_i & y_i & \end{array} \\ \begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \end{array} \qquad \vec{c} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \qquad \vec{l} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

$$X\vec{c} = \vec{l}$$

↳

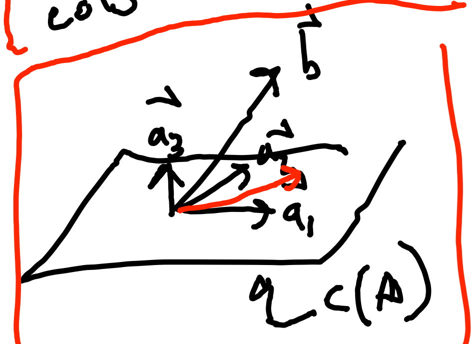
$\hat{\vec{c}} \Rightarrow$ estimate

Formula

$$\hat{\vec{c}} = (X^T X)^{-1} X^T \vec{l}$$

Condition when formula works

X has linearly independent cols $\Rightarrow (X^T X)^{-1}$ exists



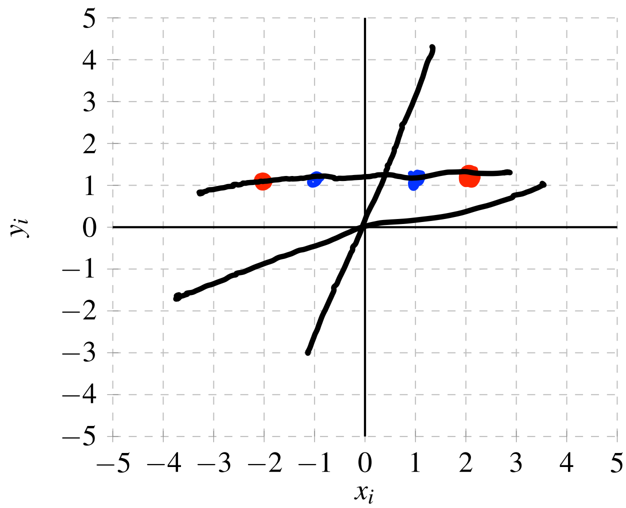
↳ Geometric intuition for why linear independence matters

Not solvable
(No unique solution to $X^T X \hat{\vec{c}} = X^T \vec{l}$, $(X^T X)^{-1}$ doesn't exist)

(b) (3 points) **Plot** the data points in the plot below with axes (x_i, y_i). **Is there a straight line such that the data points with a $+1$ label are on one side and data points with a -1 label are on the other side? Answer yes or no, and if yes, draw the line.**

x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 2: *
Table repeated for your convenience: Labels for data you are classifying



No line (straight) separates the points of different categories

(c) (6 points) You now consider a model with a quadratic term: $l_i \approx \alpha x_i + \beta x_i^2$ with $\alpha, \beta \in \mathbb{R}$. *Read the equation carefully!*

Set up a least squares problem to fit the model to the data. If this problem is solvable, solve it, i.e, find the best values for α, β. If it is not solvable, justify why.

x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 3: *
Table repeated for your convenience: Labels for data you are classifying

$$\vec{l} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \qquad X = \begin{array}{c} \begin{array}{cc} x_i & x_i^2 \end{array} \\ \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \end{array} \qquad \vec{c} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

lin. indep. cols

Solve: $X\vec{c} \approx \vec{b}$

$$\hat{\vec{c}} = (X^T X)^{-1} X^T \vec{l} = \dots$$ can compute (look @ answers) not computed here

Computation by hand:

$$\hat{\vec{c}} = \left(\begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix}^T \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \right)^{-1} \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix}^T \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 10 & 0 \\ 0 & 34 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ -6 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & \frac{1}{34} \end{bmatrix} \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -\frac{3}{17} \end{bmatrix}$$

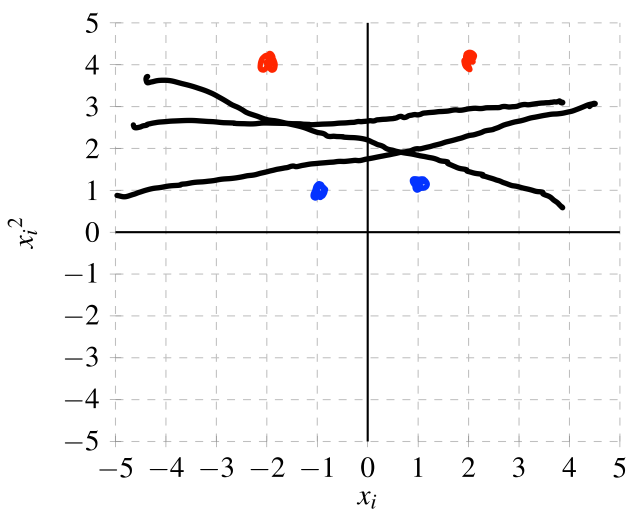
(d) (3 points) **Plot** the data points in the plot below with axes (x_i, x_i^2). **Is there a straight line such that the data points with a $+1$ label are on one side and data points with a -1 label are on the other side? Answer yes or no, and if yes, draw the line.**

x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

x_i^2
(-1, 1)
(1, 1)

x_i^2
(-2, 4)
(2, 4)

Table 4: *
Table repeated for your convenience: Labels for data you are classifying



now separable by a line. many choices

Extra:

ML Type perspective

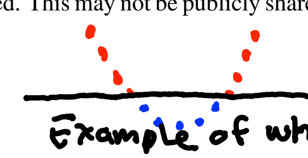
Imagine your data is on a line:

Can't separate red and blue! (with a straight line)

Can separate with a single line

Maybe we're looking at it in a way that doesn't help us (choice of model).

what if we can change the shape by using a different model to gain separability by a line?



Cast into 2-D! Now separable by a line!

Example of why model choice is important

(e) (4 points) Finally you consider the model: $l_i \approx \alpha x_i + \beta x_i^2 + \gamma$, where $\alpha, \beta, \gamma \in \mathbb{R}$. Independent of the work you have done so far, **would you expect this model or the model in part (c) (i.e. $l_i \approx \alpha x_i + \beta x_i^2$) to have a smaller error in fitting the data? Explain why.**

x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 5: *
Table repeated for your convenience: Labels for data you are classifying

Smaller error than the previous model in (d).

We can chip more off of $\vec{l}$ with another column/direction $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$

$$\begin{bmatrix} x_1 & x_1^2 & 1 \\ x_2 & x_2^2 & 1 \\ x_3 & x_3^2 & 1 \\ x_4 & x_4^2 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \approx \vec{l}$$

some part of $\vec{l}$ may lie in the direction of $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$