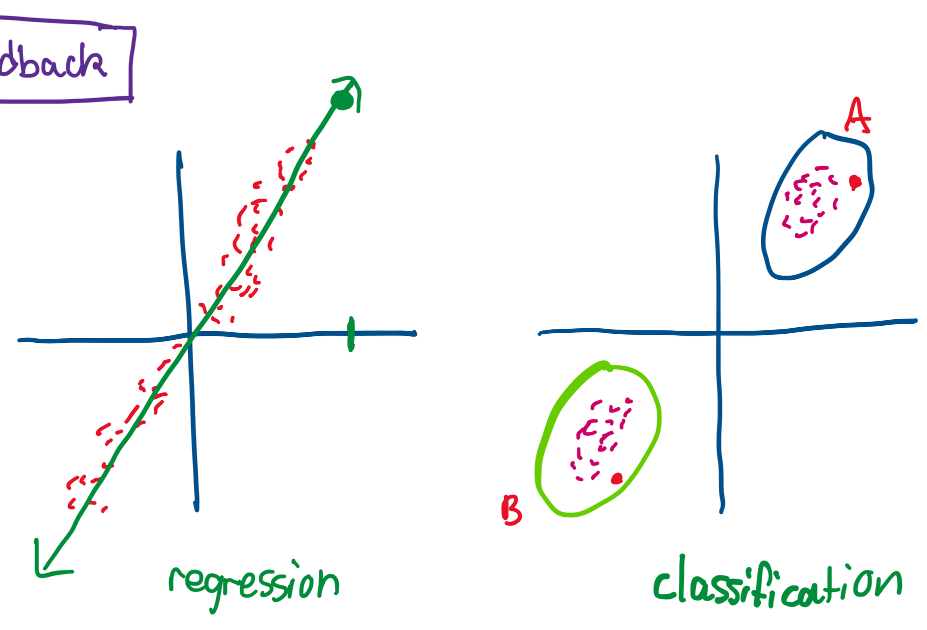
EECS 16A Dis 14A Notes

Sunday, November 28, 2021 6:24 PM

Feedback form: tinyurl.com/anusha16afeedback

Nul(A): $A\vec{x} = \vec{0}$

Nul(A) = $\{\vec{0}\}$



1. Building a classifier

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) 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.

$\vec{x} = (A^TA)^{-1} \cdot A^Tb$

$$\underbrace{A = \begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}}_{\vec{x}} \approx \underbrace{\begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix}}_{b}$$

$-2 \cdot \alpha + 1 \cdot \beta + \gamma = l_1 = -1$

$(A^TA)^{-1}$

A^TA to have LI cols

A has LD cols so Nul(A) is non-trivial

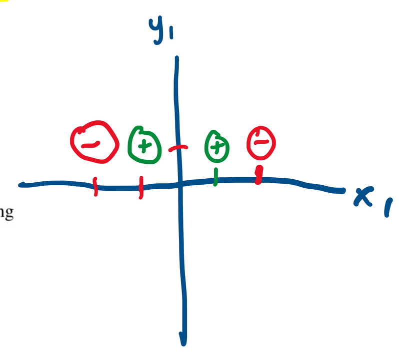
Nul(A) = Nul(A^TA) → not trivial

NOT SOLVABLE (can't find $(A^TA)^{-1}$ since Nul(A^TA) is non-trivial)

(b) 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: *
Labels for data you are classifying



NO STRAIGHT LINE

(c) 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!*

$l_1 \approx \alpha \cdot x_1 \cdot \beta \cdot y_1 + \gamma$

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: *
Labels for data you are classifying

$$A = \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

$x_i \quad x_i^2$

Nul(A^TA) = $\{\vec{0}\}$

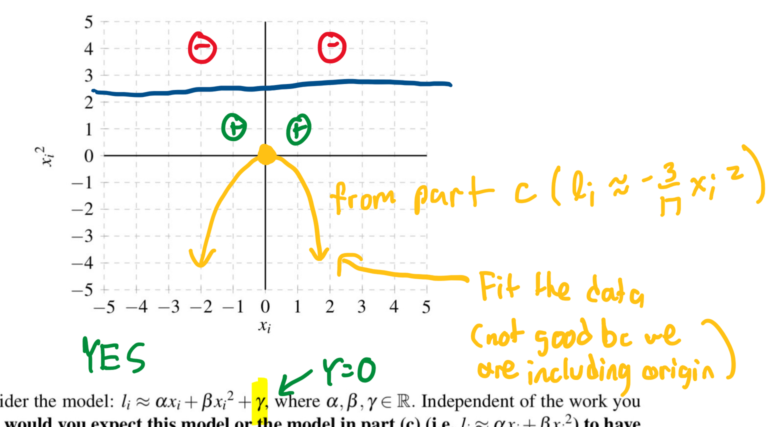
$(A^TA)^{-1} \cdot A^Tb = \begin{bmatrix} 0 & -\frac{3}{17} \end{bmatrix}^T = \begin{bmatrix} 0 \\ -\frac{3}{17} \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

Model: $l_i \approx -\frac{3}{17} x_i^2$

(d) 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

Table 4: *
Labels for data you are classifying



YES

$\gamma = 0$

(e) 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.**

Better model because of γ (more freedom)

As seen in the graph for 1c, the parabola does not fit the data well when we constrain it to include origin

2. Least Squares with Orthogonal Columns

(a) Consider a least squares problem of the form

$$\min_{\vec{x}} \left\| \vec{b} - \mathbf{A}\vec{x} \right\|^2 = \min_{\vec{x}} \left\| \mathbf{A}\vec{x} - \vec{b} \right\|^2 = \min_{\vec{x}} \left\| \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} - \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2$$

Let the solution be $\hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$.

Label the following elements in the diagram below.

$\cancel{\text{span}\{\vec{a}_1, \vec{a}_2\}}$ (1), $\cancel{\vec{e} = \vec{b} - \mathbf{A}\hat{\vec{x}}}$ (2), $\cancel{\mathbf{A}\hat{\vec{x}}}$ (3), $\vec{a}_1\hat{x}_1$ (4), $\vec{a}_2\hat{x}_2$ (5), $\cancel{\text{colspace}(\mathbf{A})}$ (6)

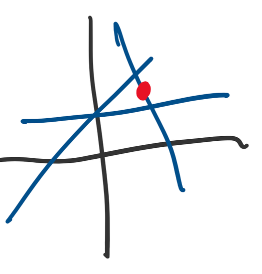
$A\vec{x} = \vec{b}$ $\quad \hat{\vec{x}} = A \cdot (A^TA)^{-1} \cdot A^Tb$

$A\hat{\vec{x}} = \hat{b}$ closest to b in Col(A)

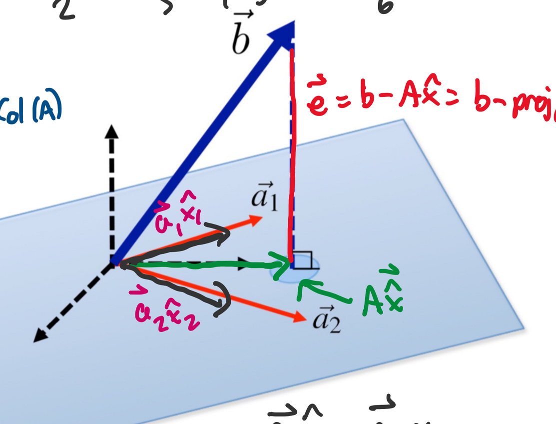
Least Squares: $\vec{x} = A(A^TA)^{-1}A^Tb$

$A\vec{x} = \vec{b}$

(1) ∞ sol
(2) unique sol
(3) no sol



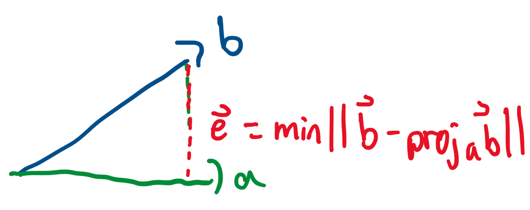
$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}$ = Col(A)



span$\{a_1, a_2\}$ = Col(A)

$\vec{a}_1 \hat{x}_1$, $\vec{a}_2 \hat{x}_2$

$$A\hat{x} = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix} \cdot \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \vec{a}_1 \hat{x}_1 + \vec{a}_2 \hat{x}_2 = \hat{b}$$



$A\vec{x} = \vec{b}$

$\vec{b} \notin$ Col(A)

$\vec{b} = \hat{b} + \vec{e}$

$\hat{b} \in$ Col(A), $\vec{e}$ min

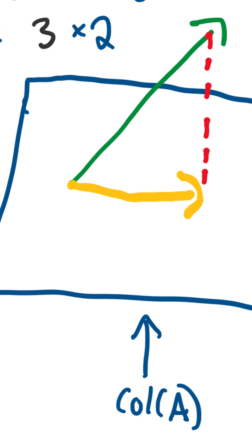
$A\hat{\vec{x}} = \hat{b}$

$A\hat{\vec{x}} \approx \vec{b}$

$A \in (3, 2)$

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} b \end{bmatrix}$$

$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix} = 3 \times 2$ $\quad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

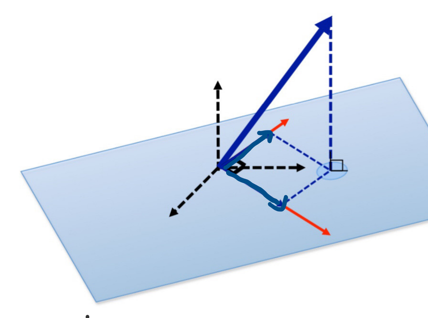


Col(A)

(b) We now consider the special case of least squares where the columns of $\mathbf{A}$ are orthogonal (illustrated in the figure below). Given that $\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$ and $A\hat{\vec{x}} = \text{proj}_{\mathbf{A}}(\vec{b}) = \hat{x}_1\vec{a}_1 + \hat{x}_2\vec{a}_2$, show that

$$\text{proj}_{\vec{a}_1}(\vec{b}) = \hat{x}_1\vec{a}_1$$
$$\text{proj}_{\vec{a}_2}(\vec{b}) = \hat{x}_2\vec{a}_2$$

$\text{proj}_{Col(A)} b = \text{proj}_{a_1} b + \text{proj}_{a_2} b$



$A = \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix}$

$\hat{\vec{x}} = (A^TA)^{-1} \cdot A^Tb \qquad A\hat{\vec{x}} = A \cdot (A^TA)^{-1} \cdot A^Tb$

$$= \underbrace{\left(\begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix} \right)^{-1}}_{} \cdot \begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \vec{b}$$

$2 \times n \quad n \times 2$

$a_1^T \cdot a_2$

inner product of orthogonal vectors is 0

$$= \left(\begin{bmatrix} a_1^Ta_1 & \cancel{a_1^Ta_2} \to 0 \\ \cancel{a_2^Ta_1} \to 0 & a_2^Ta_2 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \vec{b}$$

$$= \left(\begin{bmatrix} a_1^Ta_1 & 0 \\ 0 & a_2^Ta_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \cdot \vec{b}$$

$\|a\| = \sqrt{\langle a, a \rangle} = \sqrt{a^Ta}$

$$= \left(\begin{bmatrix} \|a_1\|^2 & 0 \\ 0 & \|a_2\|^2 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \cdot \vec{b} = \underbrace{\begin{bmatrix} \frac{1}{\|a_1\|^2} & 0 \\ 0 & \frac{1}{\|a_2\|^2} \end{bmatrix}}_{(A^TA)^{-1}} \cdot \underbrace{\begin{bmatrix} — a_1^T — \\ — a_2^T — \end{bmatrix} \cdot \vec{b}}_{A^Tb}$$

$2 \times 2 \quad 2 \times n \quad n \times 1$

$$= \begin{bmatrix} \frac{a_1^T\vec{b}}{\|a_1\|^2} \\ \frac{a_2^T\vec{b}}{\|a_2\|^2} \end{bmatrix}$$

$\text{scalproj}_{a_1} b = \frac{a_1^Tb}{\|a_1\|^2}$

$\text{scalproj}_{a_2} b = \frac{a_2^T \cdot b}{\|a_2\|^2}$

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} \text{scal proj}_{a_1} \vec{b} \\ \text{scal proj}_{a_2} \vec{b} \end{bmatrix}$$

Least squares solution when Columns of A are orthogonal

Scalar projection (i.e $\frac{a_2^Tb}{\|a_2\|^2}$ instead of $\frac{a_2^Tb}{\|a_2\|^2} a_2$

(c) Compute the least squares solution to

$$\min_{\vec{x}} \left\| \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|^2.$$

$\vec{b} - A \ \vec{x}$

$\min \|\vec{b} - A \cdot \vec{x}\|$

$$\begin{bmatrix} \frac{a_1^T\vec{b}}{\|a_1\|^2} \\ \frac{a_2^T\vec{b}}{\|a_2\|^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{1} \\ \frac{3}{1} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$$

$\|a_1\|^2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = 1$

$\|a_2\|^2 = 1$