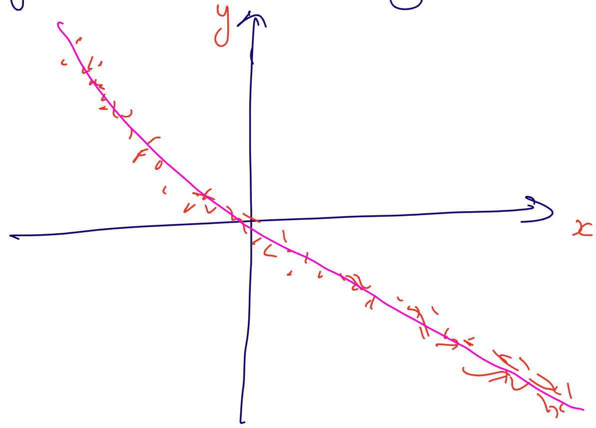
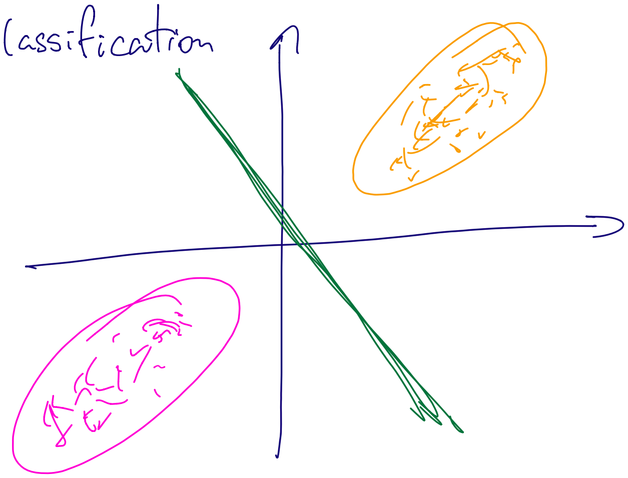
Regression + Modelling



$x \to \boxed{f(x)} \to y$

Find $f(x)$

Classification



$x \to \boxed{} \to y$

Describe grouping or differentiation

More in EECS16B

EECS 16A Designing Information Devices and Systems I
Spring 2021 Discussion 13B

$(x_i, y_i) \rightarrow \boxed{} \rightarrow l_i$

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 $l_i \in \{-1, 1\}$.

$y = 1\cdot\alpha_0 + x\cdot\alpha_1 + x^2\alpha_2 + \ldots$

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

$-1\alpha + 1\beta + 1\gamma = 1$

$\vdots$

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

$$\begin{bmatrix} | & | & | \\ x_i & y_i & 1 \\ | & | & | \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} | \\ l_i \\ | \end{bmatrix}$$

$$\begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

(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 $l_i \approx \alpha x_i + \beta y_i + \gamma$. $(A^TA)^{-1}A^Tb = \hat{x}$ ↳ not LI. can't take $(A^TA)^{-1}$, can't do L.S.

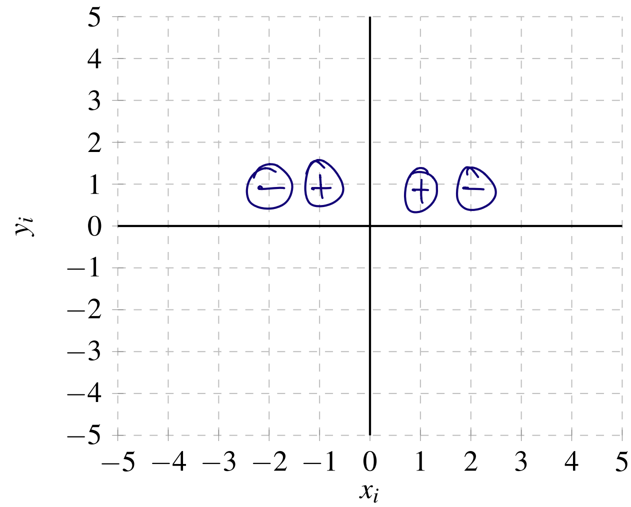
Set up a least squares problem to solve for α, β and γ. If this problem is solvable, solve it, i.e. find the best values for α, β, γ. If it is not solvable, justify why. ↳ A must have LI columns.

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

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



cannot separate with just one 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!*

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. Yes solvable.

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = (A^TA)^{-1}A^T\vec{b}$$

$$\begin{array}{cc} x_i & x_i^2 \\ \end{array}$$
$$\begin{bmatrix}-2 & 4\\ -1 & 1\\ 1 & 1\\ 2 & 4\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = \begin{bmatrix}-1\\ 1\\ 1\\ -1\end{bmatrix}_{l_i}$$

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

$$A^TA = \begin{bmatrix}-2 & -1 & 1 & 2\\ 4 & 1 & 1 & 4\end{bmatrix}\begin{bmatrix}-2 & 4\\ -1 & 1\\ 1 & 1\\ 2 & 4\end{bmatrix} = \begin{bmatrix}10 & 0\\ 0 & 34\end{bmatrix}$$

$$(A^TA)^{-1} = \begin{bmatrix}\frac{1}{10} & 0\\ 0 & \frac{1}{34}\end{bmatrix} \quad A^Tb = \begin{bmatrix}-2 & -1 & 1 & 2\\ 4 & 1 & 1 & 4\end{bmatrix}\begin{bmatrix}-1\\ 1\\ 1\\ -1\end{bmatrix} = \begin{bmatrix}0\\ -6\end{bmatrix}$$

$$(A^TA)^{-1}A^Tb = \begin{bmatrix}\frac{1}{10} & 0\\ 0 & \frac{1}{34}\end{bmatrix}\begin{bmatrix}0\\ -6\end{bmatrix} = \begin{bmatrix}0\\ \frac{-3}{17}\end{bmatrix}$$

$$\boxed{\begin{array}{l}\alpha = 0\\ \beta = \frac{-3}{17}\end{array}}$$

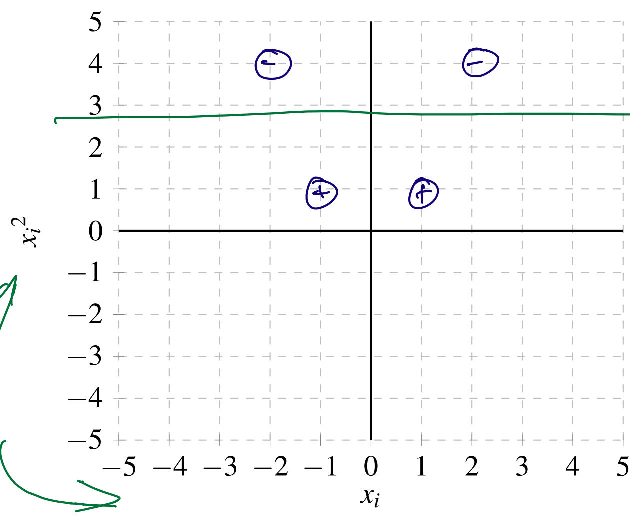
(d) Plot the data points in the plot below with axes (x_i, x_i^2). **Is there a straight line such that the data points with a $+1$ label are on one side and data points with a -1 label are on the other side? Answer yes or no, and if yes, draw the line.**

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

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

$l_i' = -\frac{3}{17} x_i^2$

x_i	x_i^2	l_i
-2	4	-1
-1	1	+1
1	1	+1
2	4	-1



some line $x_i^2 = (1-4)$

$l_i = a x_i + \beta x_i^2$

choice of axes/basis matters (more in 16B)

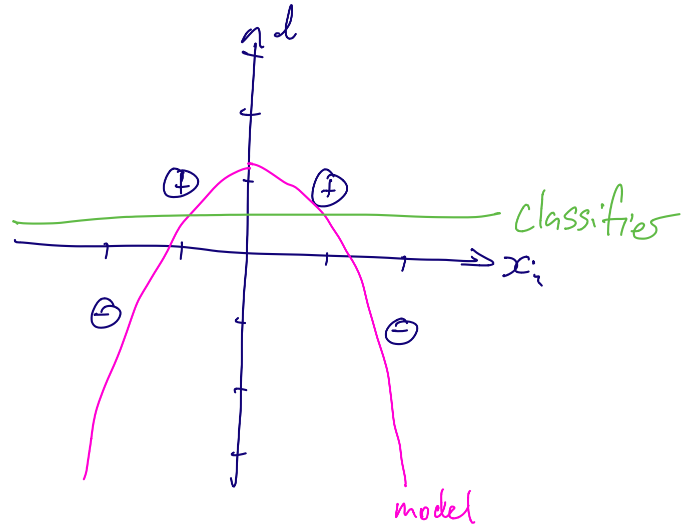
(e) Finally you consider the model: $l_i \approx \alpha x_i + \beta x_i^2 + \gamma$, where $\alpha, \beta, \gamma \in \mathbb{R}$. Independent of the work you have done so far, **would you expect this model or the model in part (c) (i.e. $l_i \approx \alpha x_i + \beta x_i^2$) to have a smaller error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

γ lets us remove the $(0,0)$ restriction.

$a = 0, \ \beta = -4/3, \ \gamma = \frac{4}{3}$

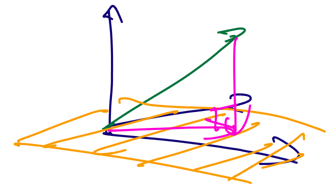
$x_i \rightarrow \boxed{\beta, \gamma} \rightarrow l_{m,i}$

x_i	l_i	$l_{m,i}$
-2	-1	-4/3
-1	+1	2/3
1	+1	2/3
2	-1	-4/3



Takeaway: we can use existing data to train and develop our model. We can use the model to make predictions about future data. How we visual or even pick our

model can hugely affect the quality of our models and predictions.



2. Orthonormal Matrices and Projections

An orthonormal matrix, $\mathbf{A}$, is a matrix whose columns, $\vec{a}_i$, are:

- Orthogonal (ie. $\langle \vec{a}_i, \vec{a}_j \rangle = 0$ when $i \neq j$)
- Normalized (ie. vectors with length equal to 1, $\|\vec{a}_i\| = 1$). This implies that $\|\vec{a}_i\|^2 = \langle \vec{a}_i, \vec{a}_i \rangle = 1$.

(a) Suppose that the matrix $\mathbf{A} \in \mathbb{R}^{N\times M}$ has linearly independent columns. The vector $\vec{y}$ in $\mathbb{R}^N$ is not in the subspace spanned by the columns of $\mathbf{A}$. What is the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$?

$\vec{y} \notin \text{Col}(A)$

LS: $\hat{x} = (A^TA)^{-1}A^T\vec{y}$

$\hat{y} = A\hat{x} = \underbrace{A(A^TA)^{-1}A^T\vec{y}}_{\text{proj}_A(y)}$

(b) Show if $\mathbf{A} \in \mathbb{R}^{N\times N}$ is an orthonormal matrix then the columns, $\vec{a}_i$, form a basis for $\mathbb{R}^N$.

see extra page

→ does not work for short/wide matrices

(c) When $\mathbf{A} \in \mathbb{R}^{N\times M}$ and $N \geq M$ (i.e. tall matrices), show that if the matrix is orthonormal, then $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

$$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_m \\ | & & | \end{bmatrix}$$

$$A^TA = \underbrace{\begin{bmatrix} — a_1^T — \\ \vdots \\ — a_m^T — \end{bmatrix}}_{M\times N} \underbrace{\begin{bmatrix} | & & | \\ a_1 & \cdots & a_m \\ | & & | \end{bmatrix}}_{N\times M} = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots \\ a_2^Ta_1 & a_2^Ta_2 & \\ a_3^Ta_1 & & \ddots \\ \vdots & & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & & 1 & \\ & & & & \ddots \end{bmatrix} = I_{M\times M}$$

note the dimensions of A^T and A

(d) Again, suppose $\mathbf{A} \in \mathbb{R}^{N\times M}$ where $N \geq M$ is an orthonormal matrix. Show that the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$ is now $\mathbf{A}\mathbf{A}^T\vec{y}$.

$$\hat{y} = A(A^TA)^{-1}A^Ty = A(I_{m\times m})^{-1}A^Ty = A(I_{m\times m})A^Ty$$

$$\boxed{\hat{y} = AA^Ty} \rightsquigarrow \hat{x} = A^Ty$$

(e) Given $\mathbf{A} \in \mathbb{R}^{N\times M} = \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and the columns of $\mathbf{A}$ are orthonormal, find the least squares solution to $\mathbf{A}\hat{\vec{x}} = \vec{y}$ where $\vec{y} = \begin{bmatrix} 5 & 12 & 7 & 8 \end{bmatrix}^T$.

$$\hat{x} = (A^TA)^{-1}A^Ty = A^Ty = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ \frac{17\sqrt{2}}{2} \end{bmatrix}$$

(2) b) Show that for orthonormal matrix A, the columns a_i form a basis for $\mathbb{R}^n$

Basis for a N-dim. subspace:
(1) N LI vectors.
(2) vectors must span subspace

Need to show these 2 properties

(1) Show vectors are LI.

from defn. of LI

Iff the vectors are LI, then $A\vec{\beta} = \vec{0}$ iff $\vec{\beta} = \vec{0}$

↳ Show that if $A\vec{\beta} = \vec{0}$, then $\vec{\beta} = \vec{0}$

$$A\vec{\beta} = (\beta_1 a_1 + \beta_2 a_2 + \dots) = \vec{0}$$

$$\langle a_i, A\vec{\beta} \rangle = (\beta_1 a_i^T a_1 + \dots + \beta_N a_i^T a_N) = \beta_i a_i^T a_i$$

$$= \langle a_i, \vec{0} \rangle = 0 = \beta_i \underbrace{a_i^T a_i}_{\text{must be } \neq 0 \text{ if } a_i \neq \vec{0}}$$

$\langle a_i, A\vec{\beta} \rangle = \langle a_i, \vec{0} \rangle$

inner product with $\vec{0}$ is always 0

$$\Downarrow$$

$$\beta_i = 0$$

$$\Rightarrow \vec{\beta} = \vec{0} \Rightarrow a_i \text{ are LI}$$

We found that $A\vec{\beta} = \vec{0}$ only for $\vec{\beta} = 0$. This only happens when a_i are LI.

(2) Show that the vectors span $\mathbb{R}^n$.

Iff col(A) spans $\mathbb{R}^n$, then for any vector $\vec{b} \in \mathbb{R}^n$, there must exist $\vec{x}$ such that $A\vec{x} = \vec{b}$.

It is sufficient to show that $\vec{x}$ always exists.

a_i are LI $\Rightarrow A^{-1}$ exist $\Rightarrow \vec{x} = A^{-1}\vec{b}$

$\vec{x}$ exists for any vector $\vec{b}$. So col(A) must span all of $\mathbb{R}^n$.