Dis 13B EECS16A

tinyurl.com/miyuki-feedback

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

$l_i \approx \alpha x_i + \beta y_i + \gamma$

$-1 \approx \alpha(-2) + \beta(1) + \gamma$

$1 \approx \alpha(-1) + \beta(1) + \gamma$

$\vdots$

(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.

Solvable?
is it possible to solve for $\hat{\vec{x}}$

$\boxed{\text{no because } A \text{ has lin dep cols and } A^TA \text{ is not invertible}}$

$$\underbrace{\begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}}_{\vec{b}} \approx \underbrace{\begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}}_{\vec{x}}$$

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

$\hat{\vec{x}}$ is only solvable if A^TA is invertible

* lin dep def:
$\alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n = \vec{0}$
if you can find some α_i not all equal to zero, then lin dep

* know A^TA is a square matrix
but not all square matrices are invertible

Ex: if $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$ A^TA is invertible

Ex: if $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$ A^TA is not invertible

* A^TA is only invertible if cols of A are linearly independent

→ to prove think about $N(A)$ and how it relates to $N(A^TA)$

→ if cols are lin ind $N(A)$ is trivial

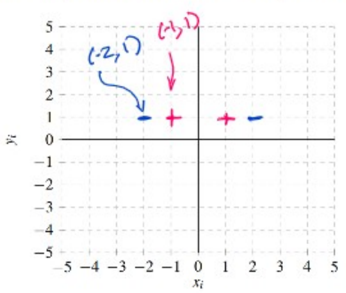
→ $N(A^TA) = N(A)$ is trivial

→ A^TA is invertible

(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
can't find a line where + are on one side and − are on the other

(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

$l_i \approx \alpha x_i + \beta x_i^2$

$-1 \approx \alpha(-2) + \beta(-2)^2$

$1 \approx \alpha(-1) + \beta(-1)^2$

$1 \approx \alpha(1) + \beta(1)^2$

$-1 \approx \alpha(2) + \beta(2)^2$

Solvable?
can we find $\hat{\vec{x}}$?

yes because cols of A are lin ind

$$\underbrace{\begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}}_{\vec{b}} \approx \underbrace{\begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\vec{x}}$$

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

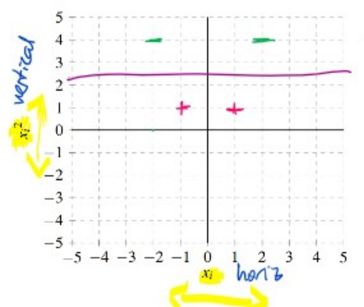
$= \begin{bmatrix} 0 \\ -\frac{3}{17} \end{bmatrix}$

(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

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

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



yes

(x_i, y_i)

(x_i, x_i^2)

(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

* error: $\|\vec{e}\|^2 = \|\vec{b} - A\hat{\vec{x}}\|^2$

$l_i \approx \alpha x_i + \beta x_i^2 + \gamma$

$$[\ ] = [\ ] \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$$

* smaller error

one more parameter to vary (γ)

* can find lines not going through origin

$l_i \approx \alpha x_i + \beta x_i^2$

$$[\ ] = [\ ] \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

* only finds lines through origin

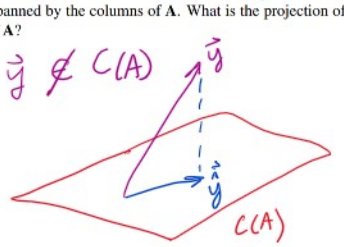
2. Orthonormal Matrices and Projections

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 $\mathbf{A}$?

$\vec{y} \notin C(A)$



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

$A\vec{x} = \vec{y} \rightarrow$ no sol

$A\hat{\vec{x}} = \hat{\vec{y}} \rightarrow$ unique sol

(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 $\mathbb{R}^N$.

Defs: orthonormal matrix $\quad A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix} \quad \langle \vec{a}_i, \vec{a}_j \rangle = 0 \quad \langle \vec{a}_i, \vec{a}_i \rangle = \|\vec{a}_i\|^2 = 1$

basis for $\mathbb{R}^N$ $\quad$ ★① need exactly N linearly independent vectors $\{\vec{v}_1, \dots, \vec{v}_N\}$

★② span $\mathbb{R}^N$ i.e. any $\vec{x} \in \mathbb{R}^N$ $\vec{x} = \alpha_i \vec{v}_1 + \cdots + \alpha_N \vec{v}_N$ for some $\alpha_1, \dots, \alpha_n$

linear independence $\quad \alpha_1 \vec{v}_1 + \cdots + \alpha_N \vec{v}_N = \vec{0}$

if $\alpha_1 = \cdots = \alpha_N = 0$ then $\{\vec{v}_1, \dots, \vec{v}_N\}$ are lin ind

★ need to show $\vec{a}_1, \dots, \vec{a}_N$ are lin ind and span $\mathbb{R}^N$

$$\beta_1 \vec{a}_1 + \cdots + \beta_n \vec{a}_N = \vec{0} \qquad \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix} \vec{\beta} = \vec{0}$$

$$\langle \vec{a}_i, \beta_1 \vec{a}_1 + \cdots + \beta_n \vec{a}_N \rangle = \langle \vec{a}_i, \vec{0} \rangle$$

$$\beta_1 \cancelto{0}{\langle \vec{a}_i, \vec{a}_1 \rangle} + \cdots + \beta_i \underbrace{\langle \vec{a}_i, \vec{a}_i \rangle}_{1} + \cdots + \beta_n \cancelto{0}{\langle \vec{a}_i, \vec{a}_N \rangle} = 0$$

$$0 + \cdots + 0 + \beta_i + 0 + \cdots + 0 = 0 \Rightarrow \beta_i = 0$$

can be shown for any $\beta_i \Rightarrow \vec{\beta} = \vec{0}$

therefore cols of A are lin ind

now we need to show that it spans $\mathbb{R}^N$

for some $\vec{x} \in \mathbb{R}^N$ $\vec{x} = A\vec{\beta} = \beta_1 \vec{a}_1 + \cdots + \beta_N \vec{a}_N \qquad \beta_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} =$

have to show we can find a unique $\vec{\beta}$

because cols of A are lin ind $\Rightarrow$ A is invertible

$\vec{\beta} = A^{-1} \vec{x}$ have a unique sol

$\Rightarrow$ cols of A span $\mathbb{R}^N$

$\Rightarrow$ cols of A are a basis for $\mathbb{R}^N$

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

more rows than cols

$$A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_M \\ | & & | \end{bmatrix} \qquad A^T = \begin{bmatrix} — & \vec{a}_1^T & — \\ & \vdots & \\ — & \vec{a}_M^T & — \end{bmatrix}$$

$$A^TA = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_M \\ | & & | \end{bmatrix}\begin{bmatrix} — & \vec{a}_1^T & — \\ & \vdots & \\ — & \vec{a}_M^T & — \end{bmatrix} = \begin{bmatrix} \underbrace{\vec{a}_1^T\vec{a}_1}_{1} & \cancel{\vec{a}_1^T\vec{a}_2}\,0 & & \cancel{\vec{a}_1^T\vec{a}_M}\,0 \\ \cancel{\vec{a}_2^T\vec{a}_1}\,0 & \underbrace{\vec{a}_2\vec{a}_2}_{1} & & \vdots \\ \vdots & \vdots & \ddots & \\ & & & \underbrace{\vec{a}_m\vec{a}_m}_{1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & & 0 \\ 0 & 1 & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & & 1 \end{bmatrix} = I_{M\times M}$$

(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 $\mathbf{A}\mathbf{A}^T\vec{y}$.

* from part c) $A^TA = I$

$$\begin{aligned} \hat{\vec{y}} &= A(A^TA)^{-1}A^T\vec{y} \\ &= A\,(I)^{-1}A^T\vec{y} \\ &= AIA^T\vec{y} \\ &= AA^T\vec{y} \end{aligned}$$

(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 $\vec{y} = \begin{bmatrix} 5 & 12 & 7 & 8 \end{bmatrix}^T$.

* remember for matrices w/ orthogonal cols

$\|\vec{a}_1\|^2 = \|\vec{a}_2\|^2 = \|\vec{a}_3\|^2 = 1$

$$\hat{\vec{x}} = (A^TA)^{-1}A^T\vec{y} = \begin{bmatrix} \frac{\langle \vec{a}_1, \vec{y}\rangle}{\|\vec{a}_1\|^2} \\ \frac{\langle \vec{a}_2, \vec{y}\rangle}{\|\vec{a}_2\|^2} \\ \frac{\langle \vec{a}_3, \vec{y}\rangle}{\|\vec{a}_3\|^2} \end{bmatrix} = \underbrace{\begin{bmatrix} \langle \vec{a}_1, \vec{y}\rangle \\ \langle \vec{a}_2, \vec{y}\rangle \\ \langle \vec{a}_3, \vec{y}\rangle \end{bmatrix}}_{\text{* for orthonormal matrices}}$$

$$= \begin{bmatrix} 8 \\ 7 \\ \frac{17\sqrt{2}}{2} \end{bmatrix}$$

* alternatively can compute using the least squares formula
$\hat{x} = (A^TA)^{-1}A^T\vec{y}$ but takes more time