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

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

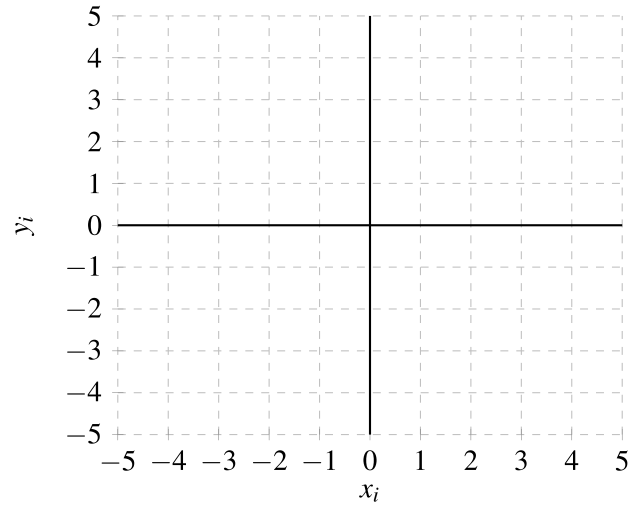
(a) You want to build a model to understand the data. You first consider a linear model, i.e. you want to find $\alpha, \beta, \gamma \in \mathbb{R}$ such that $l_i \approx \alpha x_i + \beta y_i + \gamma$.

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.

(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



(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.

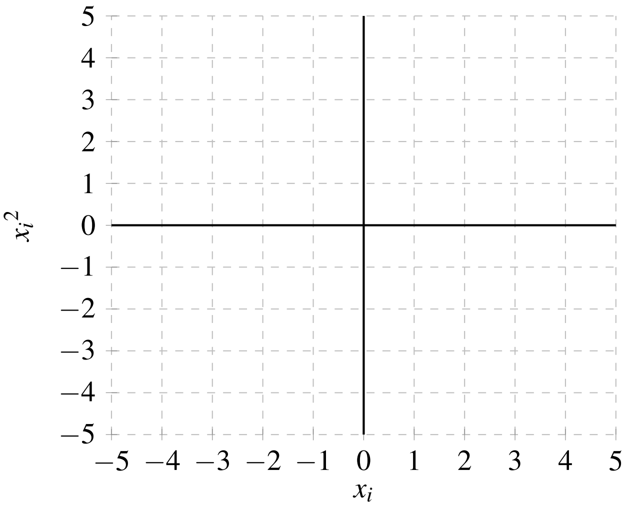
x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

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

(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



(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.**

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 $\mathbf{A}$?

Answer: When finding a projection onto a subspace, we're trying to find the "closest" vector in that subspace. This can be found by first finding $\vec{x}$ that minimizes $\|\vec{y} - \mathbf{A}\vec{x}\|$. From least squares, we know that $\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$. The projection of $\vec{y}$ onto the columns of $\mathbf{A}$ is then $\hat{\vec{y}} = \mathbf{A}\hat{\vec{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$.

(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 $\mathbb{R}^N$.

Answer:

We want to show that the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^N$. To show that the columns form a basis for $\mathbb{R}^N$ we need to show two things:

- The columns must form a set of N linearly independent vectors.
- Any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the vectors in the set.

We already know we have N vectors, so first we will show they are linearly independent. We shall do this by showing that $\mathbf{A}\vec{\beta} = \vec{0}$ implies that $\vec{\beta}$ can be only $\vec{0}$.

$$\mathbf{A}\vec{\beta} = \vec{0} \tag{1}$$

$$\beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N = \vec{0} \tag{2}$$

Then to exploit the properties of orthogonal vectors, we consider taking the inner product of each side of the above equation with $\vec{a}_i$.

$$\langle \vec{a}_i, \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \rangle = \langle \vec{a}_i, \vec{0} \rangle = 0 \tag{3}$$

Now we apply the distributive property of the inner product and the definition of orthonormal vectors,

$$\langle \vec{a}_i, \beta_1\vec{a}_1 \rangle + \ldots + \langle \vec{a}_i, \beta_i\vec{a}_i \rangle + \ldots + \langle \vec{a}_i, \beta_N\vec{a}_N \rangle = 0 \tag{4}$$

$$0 + \ldots + \beta_i\langle \vec{a}_i, \vec{a}_i \rangle + \ldots + 0 = 0 \tag{5}$$

$$0 + \ldots + \beta_i\vec{a}_i^T\vec{a}_i + \ldots + 0 = 0 \tag{6}$$

Because $\vec{a}_i^T\vec{a}_i = 1$, $\beta_i = 0$ for the equation to hold. Then, since this is true for all i from 1 to N, all the elements of the vector beta must be zero ($\vec{\beta} = \vec{0}$). Because $\vec{x} = \vec{0}$ implies $\vec{\beta} = \vec{0}$, the columns of $\mathbf{A}$ are linearly independent.

Now, we will show that any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the columns of $\mathbf{A}$.

$$\vec{x} = \mathbf{A}\vec{\beta} = \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \tag{7}$$

Because we know that the N columns of $\mathbf{A}$ are linearly independent, then there exists $\mathbf{A}^{-1}$. Applying the inverse to the equation above,

$$\mathbf{A}^{-1}\mathbf{A}\vec{\beta} = \mathbf{A}^{-1}\vec{x} \tag{8}$$

$$\vec{\beta} = \mathbf{A}^{-1}\vec{x}, \tag{9}$$

we find that there exists a unqiue β that allow us to represent any $\vec{x}$ as a linear combination of the columns of $\mathbf{A}$.

(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 $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

Answer: Want to show $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \vec{a}_1^T\vec{a}_2 & \dots & \vec{a}_1^T\vec{a}_n \\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \dots & \vec{a}_2^T\vec{a}_n \\ \vdots & \vdots & & \vdots \end{bmatrix} = \mathbf{I}_{M\times M} \tag{10}$$

When $\vec{a}_i^T\vec{a}_i = \|\vec{a}_i\|^2 = 1$ and when $i \neq j$, $\vec{a}_i^T\vec{a}_j = 0$ because the column vectors are orthogonal.

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

Answer:

Starting with the result from part (a),

$$\mathbf{A}\hat{\vec{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}, \tag{11}$$

we can apply the result from part (c),

$$\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \mathbf{A}\mathbf{I}\mathbf{A}^T\vec{y} \tag{12}$$

$$= \mathbf{A}\mathbf{A}^T\vec{y} \tag{13}$$

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

Answer:

Method 1:

Since the columns of $\mathbf{A}$ are orthonormal, from part (d) we know that

$\hat{\vec{x}} = \mathbf{A}^T\vec{y} = \begin{bmatrix} \langle \vec{a}_1, \vec{y} \rangle \\ \langle \vec{a}_2, \vec{y} \rangle \\ \langle \vec{a}_3, \vec{y} \rangle \end{bmatrix}$.

Note that this is equivalent to projecting $\vec{y}$ onto each column of $\mathbf{A}$:

$\hat{x_1} = \frac{\langle \vec{a}_1, \vec{y} \rangle}{||\vec{a}_1||^2} = \langle \vec{a}_1, \vec{y} \rangle = 8$

$\hat{x_2} = \frac{\langle \vec{a}_2, \vec{y} \rangle}{||\vec{a}_2||^2} = \langle \vec{a}_2, \vec{y} \rangle = 7$

$\hat{x_3} = \frac{\langle \vec{a}_3, \vec{y} \rangle}{||\vec{a}_3||^2} = \langle \vec{a}_3, \vec{y} \rangle = \frac{17\sqrt{2}}{2}$

Method 2 (Alternatively you can use the least squares formula):

$$\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \left(\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ \frac{17\sqrt{2}}{2} \end{bmatrix}$$