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

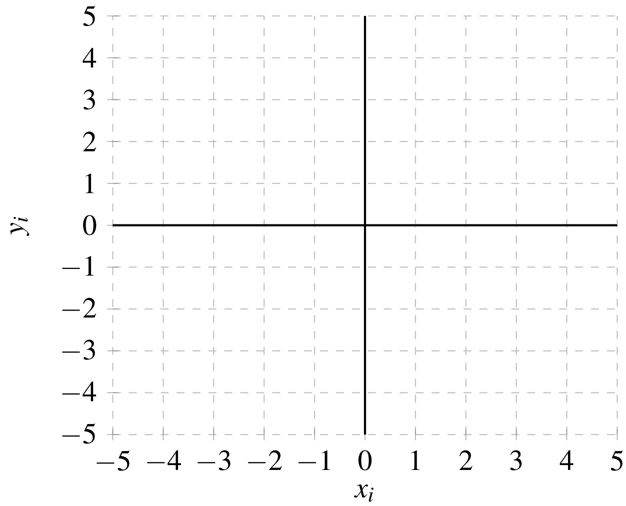
(a) You want to build a model to understand the data. You first consider a linear model, i.e. you want to find $\alpha, \beta, \gamma \in \mathbb{R}$ such that $l_i \approx \alpha x_i + \beta y_i + \gamma$.

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.

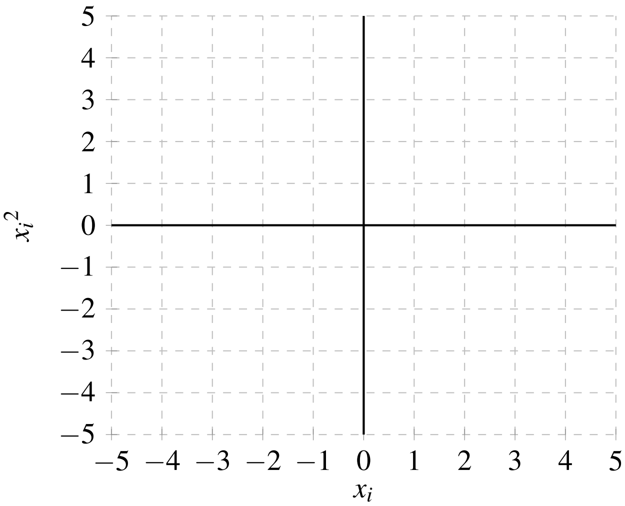
x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

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 $\mathbf{A}$?

(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 $\mathbb{R}^N$.

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

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

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