EECS 16A Designing Information Devices and Systems I
Fall 2022 Discussion 14A

1. Least Squares with Orthogonal Columns

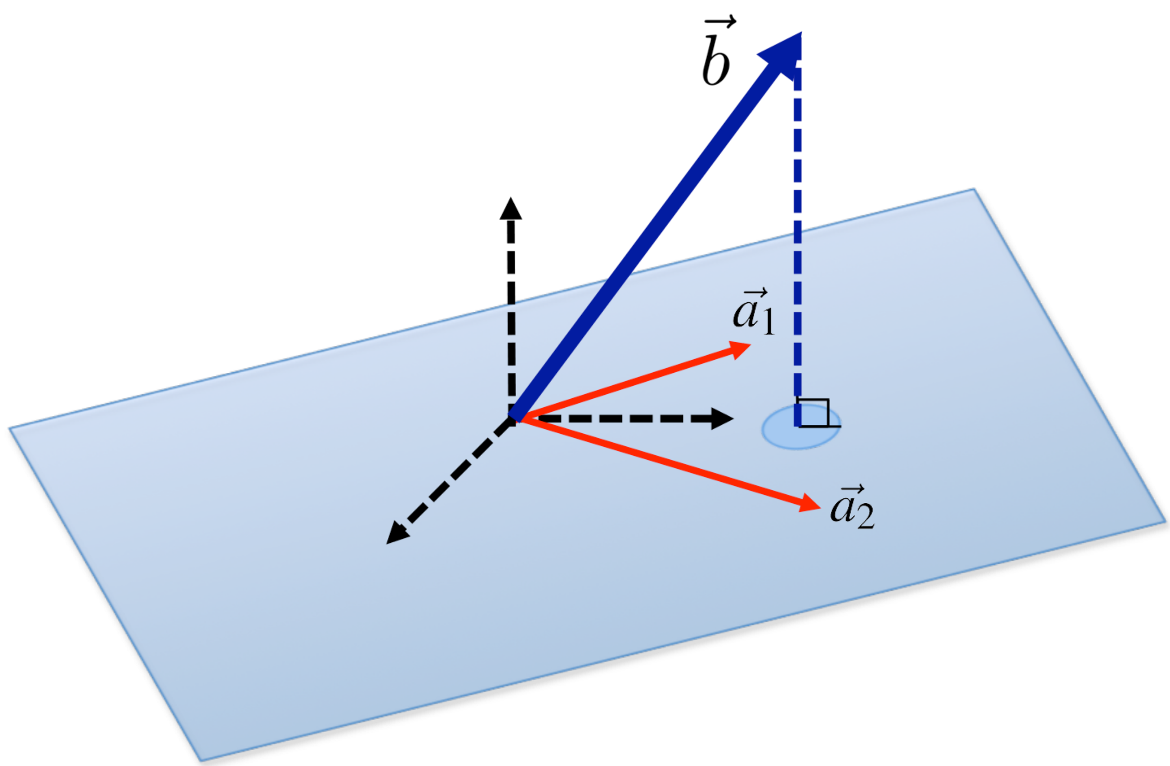
(a) Consider a least squares problem of the form

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

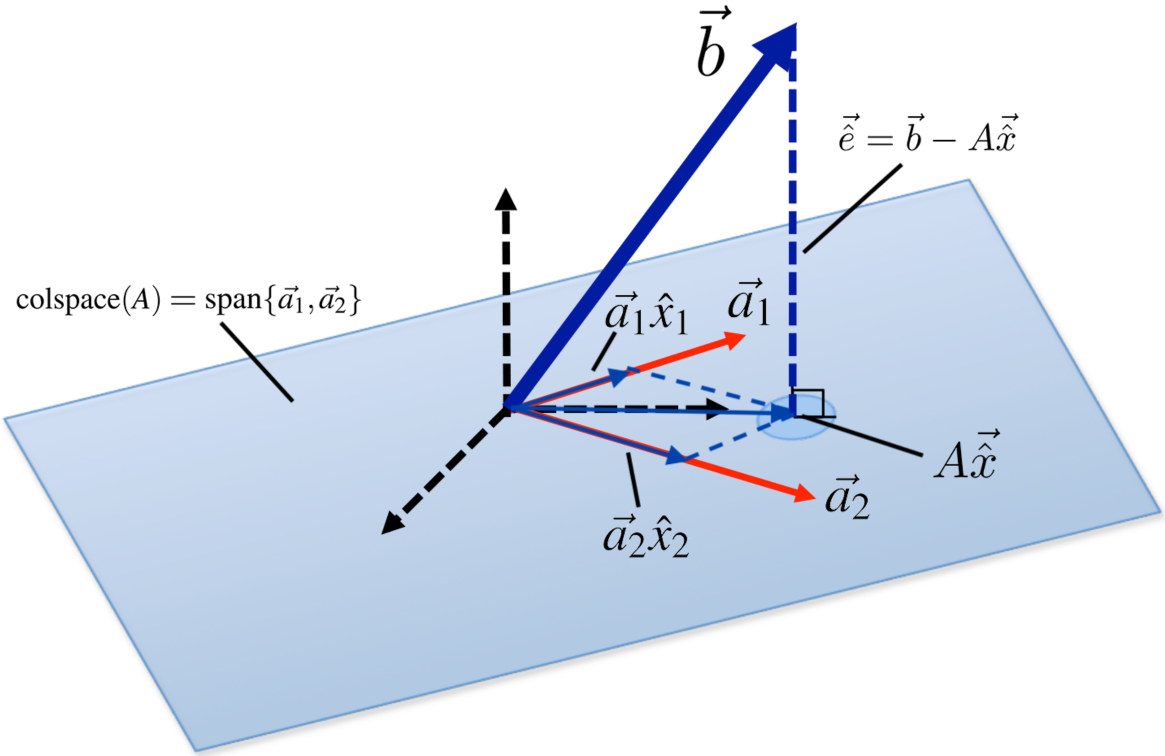
Let the solution be $\vec{\hat{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$.

Label the following elements in the diagram below.

$$\text{span}\{\vec{a}_1, \vec{a}_2\}, \quad \vec{\hat{e}} = \vec{b} - \mathbf{A}\vec{\hat{x}}, \quad \mathbf{A}\vec{\hat{x}}, \quad \vec{a}_1\hat{x}_1,\ \vec{a}_2\hat{x}_2, \quad \text{colspace}(\mathbf{A})$$



Answer:



(b) We now consider the special case of least squares where the columns of $\mathbf{A}$ are orthogonal. 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

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

Answer: The projection of $\vec{b}$ onto $\vec{a}_1$ and $\vec{a}_2$ are given by:

$$\text{proj}_{\vec{a}_1}(\vec{b}) = \frac{\langle \vec{a}_1, \vec{b}\rangle}{\|\vec{a}_1\|^2}\vec{a}_1 \qquad \text{proj}_{\vec{a}_2}(\vec{b}) = \frac{\langle \vec{a}_2, \vec{b}\rangle}{\|\vec{a}_2\|^2}\vec{a}_2$$

$$\text{Length:} \quad \frac{\langle \vec{a}_1, \vec{b}\rangle}{\|\vec{a}_1\|} \qquad \frac{\langle \vec{a}_2, \vec{b}\rangle}{\|\vec{a}_2\|}$$

The least squares solution is given by:

$$\begin{aligned}
\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} &= \left(\begin{bmatrix} - & \vec{a}_1{}^T & - \\ - & \vec{a}_2{}^T & - \end{bmatrix} \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix} \right)^{-1} \begin{bmatrix} - & \vec{a}_1{}^T & - \\ - & \vec{a}_2{}^T & - \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\|\vec{a}_1\|^2} & 0 \\ 0 & \frac{1}{\|\vec{a}_2\|^2} \end{bmatrix} \begin{bmatrix} - & \vec{a}_1{}^T & - \\ - & \vec{a}_2{}^T & - \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{\vec{a}_1{}^T\vec{b}}{\|\vec{a}_1\|^2} \\ \frac{\vec{a}_2{}^T\vec{b}}{\|\vec{a}_2\|^2} \end{bmatrix}
\end{aligned}$$

Thus,

$$\text{proj}_{\vec{a}_1}(\vec{b}) = \frac{\langle \vec{a}_1, \vec{b} \rangle}{\|\vec{a}_1\|^2}\vec{a}_1 = \frac{\vec{a}_1{}^T\vec{b}}{\|\vec{a}_1\|^2}\vec{a}_1 = \hat{x}_1\vec{a}_1$$
$$\text{proj}_{\vec{a}_2}(\vec{b}) = \frac{\langle \vec{a}_2, \vec{b} \rangle}{\|\vec{a}_2\|^2}\vec{a}_2 = \frac{\vec{a}_2{}^T\vec{b}}{\|\vec{a}_2\|^2}\vec{a}_2 = \hat{x}_2\vec{a}_2$$

(c) Compute the least squares solution to

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

Answer: Using least squares again,

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Note that the columns of $\mathbf{A}$ are orthogonal, so it is much faster to project $\vec{b}$ onto the columns of $\mathbf{A}$ than use the least squares formula to find $\vec{\hat{x}}$.

2. 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.
Solution/Answer: Rewriting the equations $\alpha x_i + \beta y_i + \gamma \approx l_i$ for $i = 1, 2, 3, 4$ in matrix form gives:

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \\ x_4 & y_4 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \approx \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

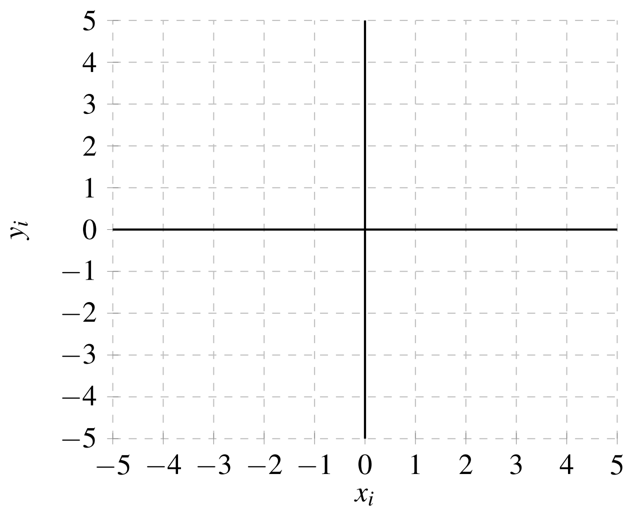
$$\mathbf{A}\vec{x} \approx \vec{b}$$

The least squares solution is $\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$. The solution only exists when the matrix $\mathbf{A}^T\mathbf{A}$ is invertible, and an equivalent condition is when all the columns of $\mathbf{A}$ are linearly independent. We see that the second and third columns of $\mathbf{A}$ are linearly dependent, so the problem is **not** solvable.

(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.

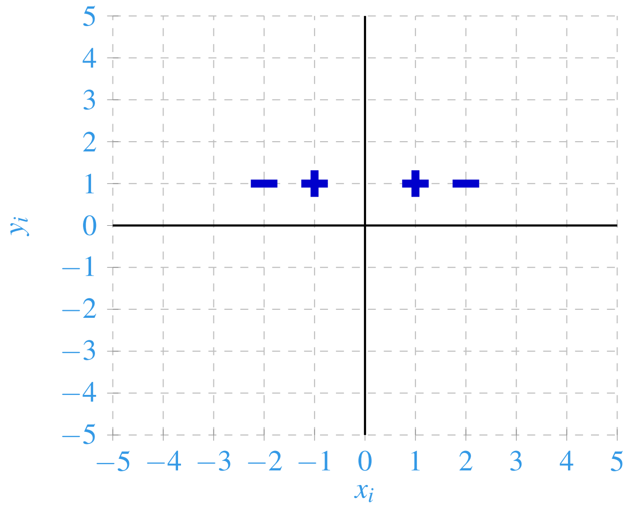
x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 2: *
Labels for data you are classifying



Solution/Answer:

The answer is no. As can be seen from the plot, the points all lie on the line $y_i = 1$, so there is no line that is able to separate the points based on their label.



(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: *
Labels for data you are classifying

Solution/Answer: Rewriting the equations $\alpha x_i + \beta x_i^2 \approx l_i$ for $i = 1,2,3,4$ in matrix form gives:

$$\underbrace{\begin{bmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \\ x_3 & x_3^2 \\ x_4 & x_4^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\mathbf{A}\vec{x} \approx \vec{b}} = \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \approx \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$$

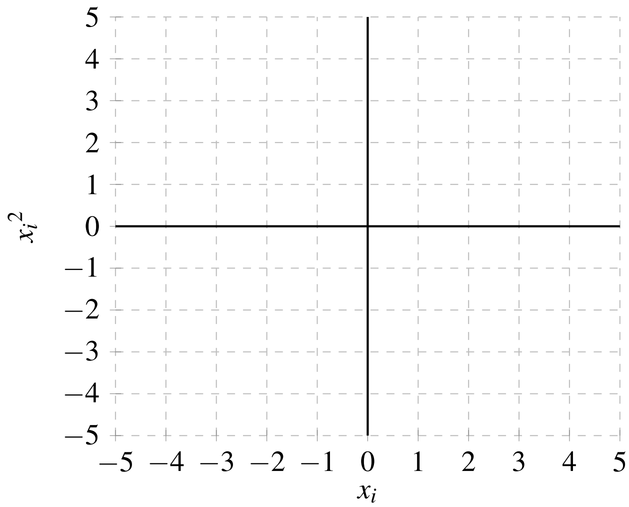
The least squares solution is $\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$. The solution only exists when the matrix $\mathbf{A}^T\mathbf{A}$ is invertible, and an equivalent condition is when all the columns of $\mathbf{A}$ are linearly independent. We see that the first and second columns of $\mathbf{A}$ are linearly independent, so the problem is solvable.

We can solve for $\hat{\vec{x}} = (\begin{bmatrix} -2 & -1 & 1 & 2 \\ 4 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} -2 & 4 \\ -1 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix})^{-1} \begin{bmatrix} -2 & -1 & 1 & 2 \\ 4 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \end{bmatrix}^T = \begin{bmatrix} 0 & \frac{-3}{17} \end{bmatrix}^T$.

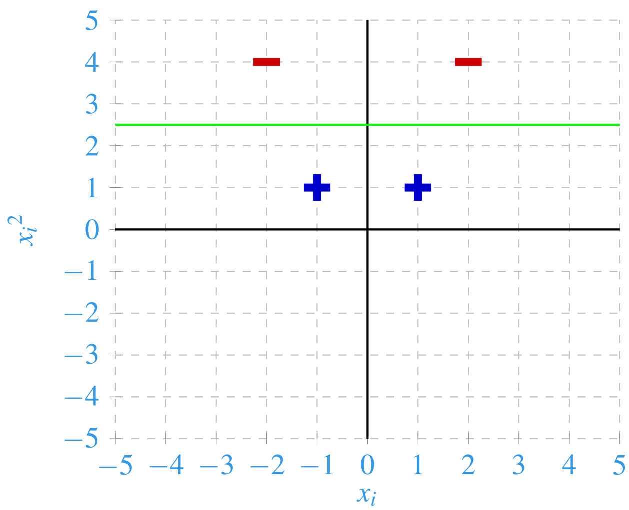
(d) Plot the data points in the plot below with axes (x_i, x_i^2). **Is there a straight line such that the data points with a $+1$ label are on one side and data points with a -1 label are on the other side? Answer yes or no, and if yes, draw the line.**

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



Solution/Answer:



The answer is yes. A line $x_i^2 = u$ where $1 < u < 4$ would separate the data points based on their labels. It is important to note that solving the least squares problem considered so far would not yield that line because the problem so far considers only lines that pass through the origin.

(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.**

Solution/Answer: We expect the model in part (e) to have a smaller error because there are more degrees of freedom. The model in part (c) only considers lines passing through the origin, while the model in part (e) considers all lines. With the model in part (e) we are able to obtain a line $x_i^2 = u$ where $1 < u < 4$ that would separate the data points based on their labels, unlike the model in part (c).

