EECS 16A Designing Information Devices and Systems I
Fall 2021 Discussion 14A

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: *
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.

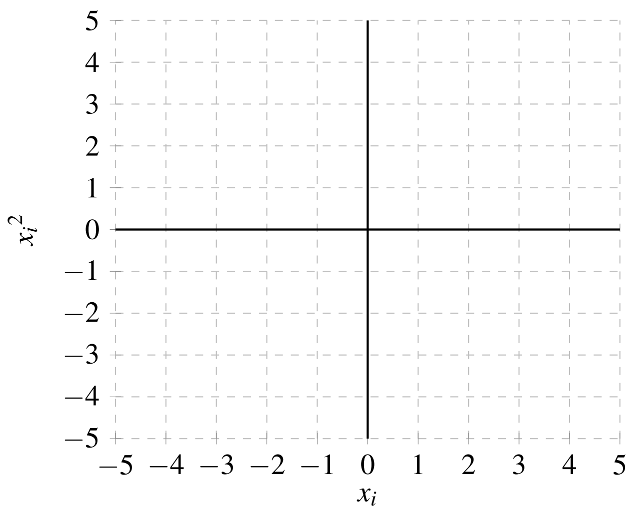
x_i	y_i	l_i
-2	1	-1
-1	1	1
1	1	1
2	1	-1

Table 3: *
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: *
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. Least Squares with Orthogonal Columns

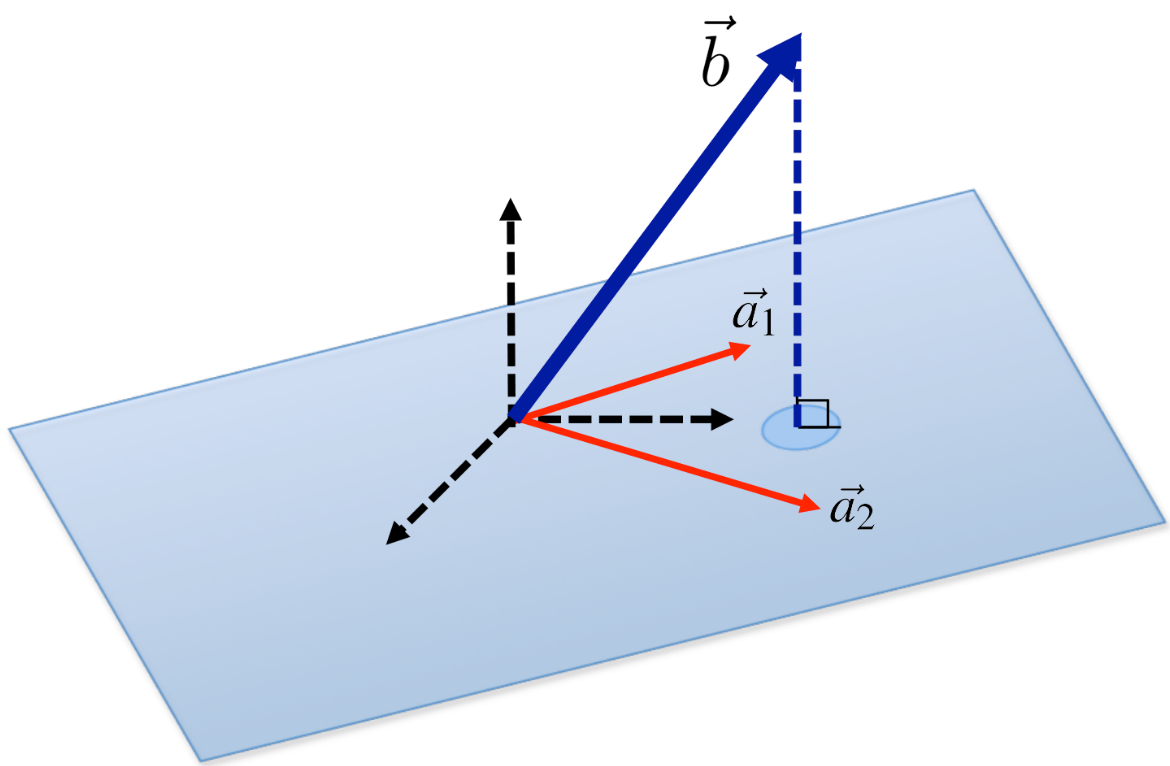
(a) Consider a least squares problem of the form

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

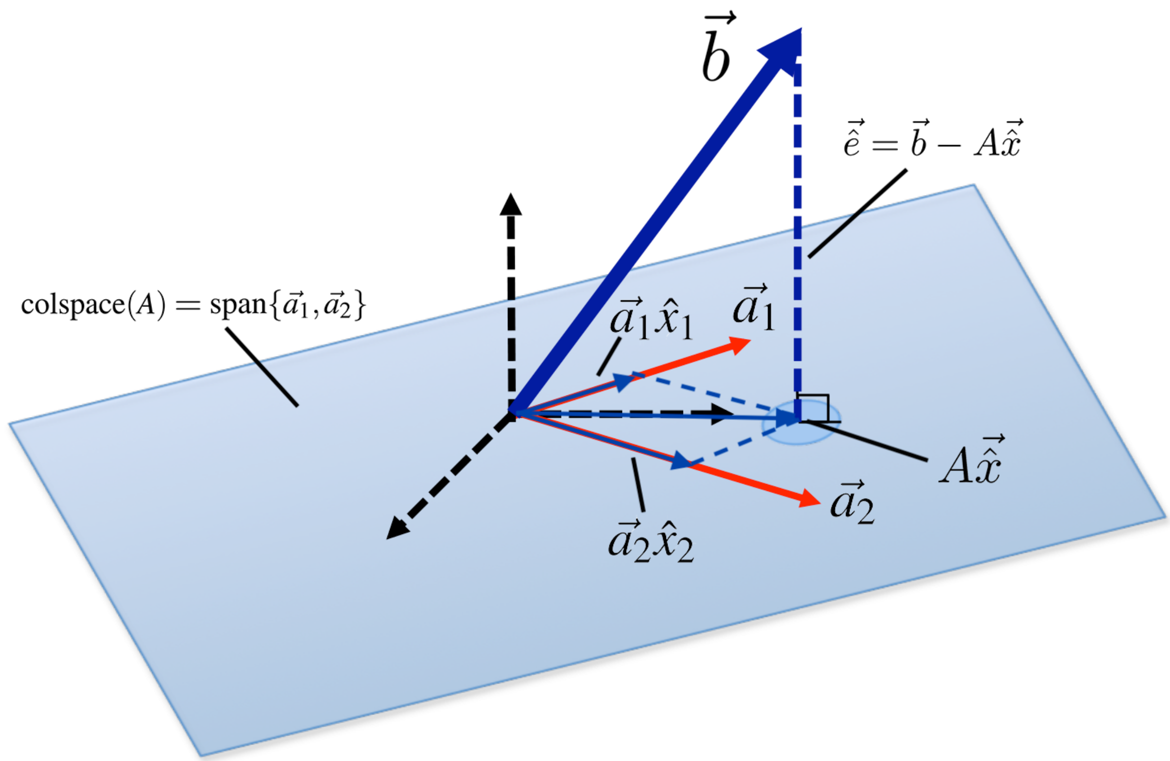
Let the solution be $\vec{\hat{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$.

Label the following elements in the diagram below.

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

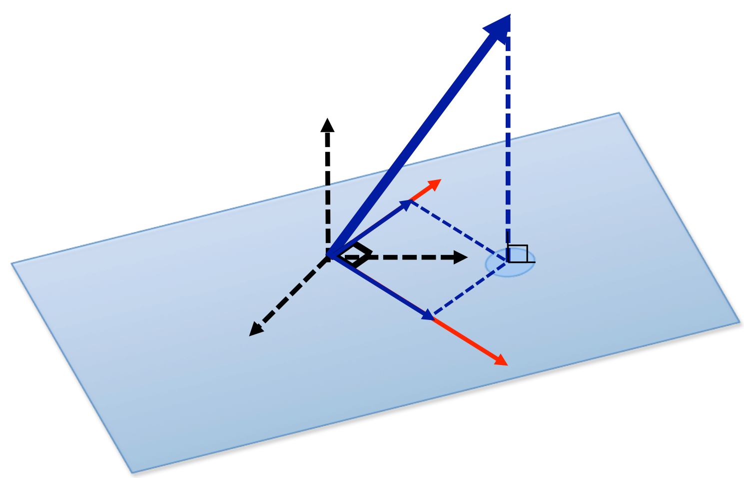


Answer:



(b) We now consider the special case of least squares where the columns of $\mathbf{A}$ are orthogonal (illustrated in the figure below). Given that $\vec{\hat{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{b}$ and $A\vec{\hat{x}} = \text{proj}_{\mathbf{A}}(\vec{b}) = \hat{x}_1\vec{a}_1 + \hat{x}_2\vec{a}_2$, 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:

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

(c) Compute the least squares solution to

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

Answer: Using least squares again,

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

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 $\hat{\vec{x}}$.

