ACM41020 Assignment 2

Linear Algebra and Regression

DUE: 10AM FRIDAY 25TH OCTOBER 2024

1. **Regression (only 12, 18, 20 graded):** Complete questions 12-22 in Section II.2 of "Linear Algebra and Learning from Data" by Gilbert Strang. The relevant pages are:

This page is devoted to the simplest and most important application of least squares: **Fitting a straight line to data.** A line $b = C + Dt$ has $n = 2$ parameters C and D. We are given $m > 2$ measurements b_i at m different times t_i. The equations $A\boldsymbol{x} = b$ (unsolvable) and $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ (solvable) are

$$A\boldsymbol{x} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \qquad A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} \widehat{C} \\ \widehat{D} \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum b_i t_i \end{bmatrix}.$$

The column space $\mathbf{C}(A)$ is a 2-dimensional plane in $\mathbf{R}^m$. The vector $\boldsymbol{b}$ is in this column space if and only if the m points (t_i, b_i) actually lie on a straight line. In that case only, $A\boldsymbol{x} = \boldsymbol{b}$ is solvable: the line is $C + Dt$. **Always $\boldsymbol{b}$ is projected to the closest $\boldsymbol{p}$ in $\mathbf{C}(A)$.**

The best line (the least squares fit) passes through the points (t_i, p_i). The error vector $\boldsymbol{e} = A\widehat{\boldsymbol{x}} - \boldsymbol{b}$ has components $b_i - p_i$. And $\boldsymbol{e}$ is perpendicular to $\boldsymbol{p}$.

There are two important ways to draw this least squares regression problem. One way shows the best line $b = \widehat{C} + \widehat{D}t$ and the errors e_i (vertical distances to the line). The second way is in $\mathbf{R}^m = m$-dimensional space. There we see the data vector $\boldsymbol{b}$, its projection $\boldsymbol{p}$ onto $\mathbf{C}(A)$, and the error vector $\boldsymbol{e}$. This is a right triangle with $||\boldsymbol{p}||^2 + ||\boldsymbol{e}||^2 = ||\boldsymbol{b}||^2$.

Problems 12 to 22 use four data points $\boldsymbol{b} = (0, 8, 8, 20)$ to bring out the key ideas.

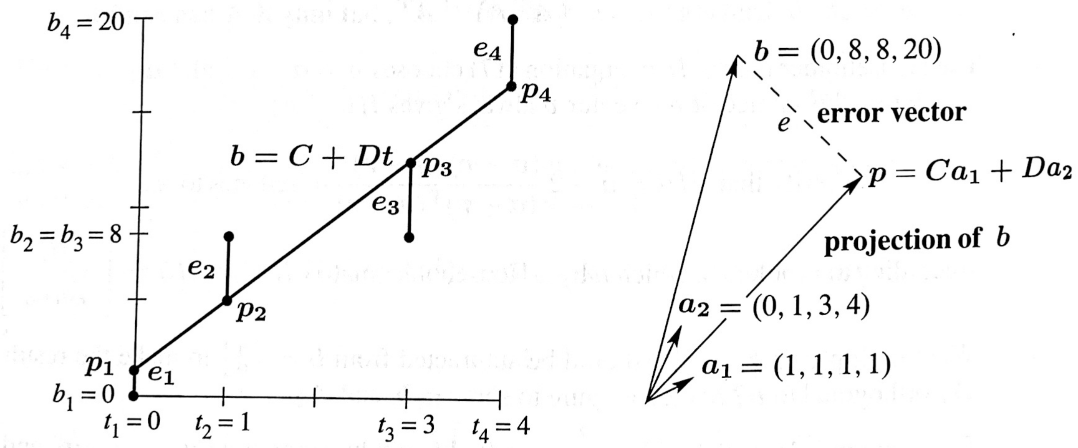


Figure II.3: The closest line $C + Dt$ in the $t - b$ plane matches $C\boldsymbol{a}_1 + D\boldsymbol{a}_2$ in $\mathbf{R}^4$.

12 With $b = 0, 8, 8, 20$ at $t = 0, 1, 3, 4$, set up and solve the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$. For the best straight line in Figure II.3a, find its four heights p_i and four errors e_i. What is the minimum squared error $E = e_1^2 + e_2^2 + e_3^2 + e_4^2$?

13 (Line $C + Dt$ does go through p's) With $b = 0, 8, 8, 20$ at times $t = 0, 1, 3, 4$, write down the four equations $A\boldsymbol{x} = \boldsymbol{b}$ (unsolvable). Change the measurements to $p = 1, 5, 13, 17$ and find an exact solution to $A\widehat{\boldsymbol{x}} = \boldsymbol{p}$.

14 Check that $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = (-1, 3, -5, 3)$ is perpendicular to both columns of the same matrix A. What is the shortest distance $\|e\|$ from $\boldsymbol{b}$ to the column space of A?

15 (By calculus) Write down $E = \|A\boldsymbol{x} - \boldsymbol{b}\|^2$ as a sum of four squares—the last one is $(C + 4D - 20)^2$. Find the derivative equations $\partial E/\partial C = 0$ and $\partial E/\partial D = 0$. Divide by 2 to obtain the normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

16 Find the height C of the best *horizontal line* to fit $\boldsymbol{b} = (0, 8, 8, 20)$. An exact fit would solve the unsolvable equations $C = 0$, $C = 8$, $C = 8$, $C = 20$. Find the 4 by 1 matrix A in these equations and solve $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. Draw the horizontal line at height $\widehat{x} = C$ and the four errors in $\boldsymbol{e}$.

17 Project $\boldsymbol{b} = (0, 8, 8, 20)$ onto the line through $\boldsymbol{a} = (1, 1, 1, 1)$. Find $\widehat{x} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ and the projection $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$. Check that $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to $\boldsymbol{a}$, and find the shortest distance $\|\boldsymbol{e}\|$ from $\boldsymbol{b}$ to the line through $\boldsymbol{a}$.

18 Find the closest line $b = Dt$, *through the origin*, to the same four points. An exact fit would solve $D \cdot 0 = 0, D \cdot 1 = 8, D \cdot 3 = 8, D \cdot 4 = 20$. Find the 4 by 1 matrix and solve $A^{\mathrm{T}}A\widehat{x} = A^{\mathrm{T}}\boldsymbol{b}$. Redraw Figure II.3a showing the best line $b = Dt$.

19 Project $\boldsymbol{b} = (0, 8, 8, 20)$ onto the line through $\boldsymbol{a} = (0, 1, 3, 4)$. Find $\widehat{x} = D$ and $\boldsymbol{p} = \widehat{x}\boldsymbol{a}$. The best C in Problem 16 and the best D in Problem 18 *do not agree* with the best $(\widehat{C}, \widehat{D})$ in Problems 11–14. That is because the two columns $(1, 1, 1, 1)$ and $(0, 1, 3, 4)$ are _____ perpendicular.

20 For the closest parabola $b = C + Dt + Et^2$ to the same four points, write down the unsolvable equations $A\boldsymbol{x} = \boldsymbol{b}$ in three unknowns $\boldsymbol{x} = (C, D, E)$. Set up the three normal equations $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ (solution not required). In Figure II.3a you are now fitting a parabola to 4 points—what is happening in Figure II.3b?

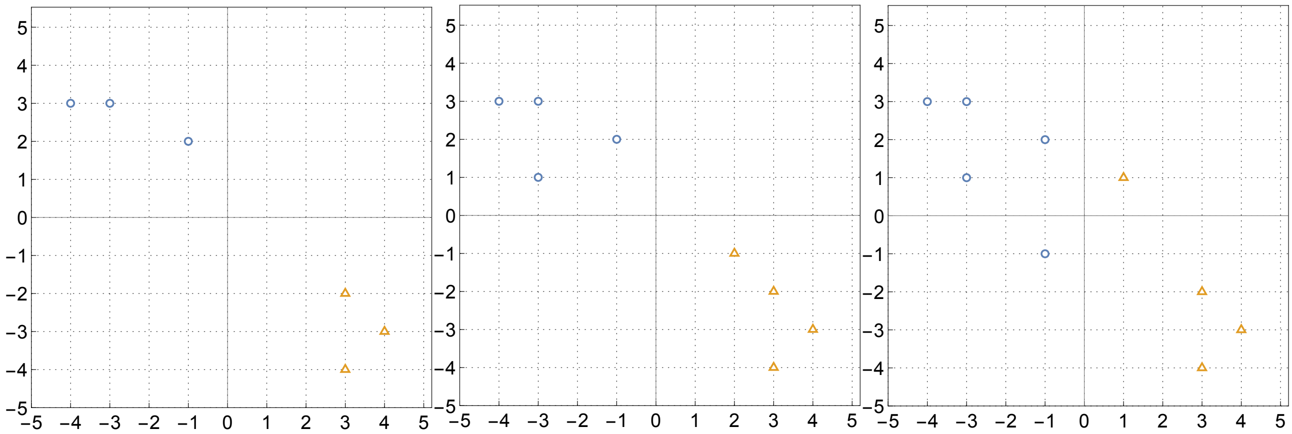
21 For the closest cubic $b = C + Dt + Et^2 + Ft^3$ to the same four points, write down the four equations $A\boldsymbol{x} = \boldsymbol{b}$. Solve them by elimination. In Figure II.3a this cubic now goes exactly through the points. What are $\boldsymbol{p}$ and $\boldsymbol{e}$?

22 The averages of the t_i and b_i are $\bar{t} = 2$ and $\bar{b} = 9$. Verify that $C + D\bar{t} = \bar{b}$. Explain !

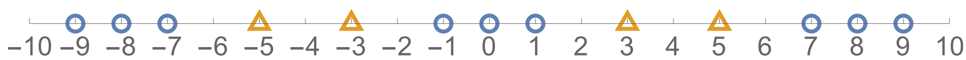
(a) Verify that the best line goes through the center point $(\bar{t}, \bar{b}) = (2, 9)$.

(b) Explain why $C + D\bar{t} = \bar{b}$ comes from the first equation in $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$.

2. **Principal Component Analysis**
   a) Using Principal Component Analysis compute a fit to the data from question 1 such that the fit minimises the **orthogonal** distance.
   b) Compute the distance from each of the points to the new fit in both the **orthogonal direction** and in the **vertical direction**.
   c) Compare these distances against the equivalent distances from the linear regression fit. Comment on the relative size of the distances in the different cases.
3. **Support Vector Machines**
   Consider the three datasets in the below graphs.
   a) Identify the support vectors.
   b) Draw the decision line and the two margins.
   c) Compute the values for the normal vector **w** and bias b defining the decision line.
   d) Compute the weights λ for each data point.



4. **Nonlinear Support Vector Machines**
   Consider the below one-dimensional data that you wish to classify with a support vector machine:



   This data is all on the same line, defined only by an x-coordinate, and it switches back and forth from clusters of orange triangles to clusters of blue circles, so it's not at all linearly separable. It is, however, possible to separate the data using the kernel function

$$K(x_1, x_2) = \cos\left(\frac{\pi}{4}x_1\right)\cos\left(\frac{\pi}{4}x_2\right) + \sin\left(\frac{\pi}{4}x_1\right)\sin\left(\frac{\pi}{4}x_2\right)$$

   a) What is the map $\phi(x)$ that projects x into the 2D space defined by this kernel?
   b) Draw the points in the new space on a 2D grid. Draw also the decision line and margins that linearly separate the points in the transformed space.

Note: The questions in this assignment can be completed with by-hand calculations, but if you wish you may use a computer to help with any number-crunching calculations.