Lecture 19

Curve Fitting

Problem: Consider the following data points in $\mathbb{R}^2$

x	y
1	0.3
2	0.7
3	1
4	1.3

We say a linear function

$$f(x) = ax + b$$

is a best fit line for this data set if it minimizes the distance between its predicted values $f(x_i)$ and the experimental values y_i,

$$\epsilon := \sqrt{\sum (f(x_i) - y_i)^2}$$ over all linear functions.

Before we find the answer let's guess.

guess $f(x) = \frac{1}{3}x$

x	$f(x)$
1	0.3333
2	0.666..
3	1
4	1.3333

What is the error (distance) for our guess

$$\epsilon := \sqrt{(0.033)^2 + \dots}$$
$$= (0.00333)^{1/2}$$

Can we do better than this?

How is this a linear algebra problem?

<u>We are trying to:</u>

Find the <u>closest</u> <u>vector</u>

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_4) \end{bmatrix} = \begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \end{bmatrix}$$

obtained by evaluating a linear function to the vector of experimental values

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.7 \\ 1 \\ 1.3 \end{bmatrix}.$$

We know how to find the closest vector in a <u>subspace</u> V to a fixed vector v (its v'').

Are vectors of the form

$$\begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \end{bmatrix}$$

obtained by evaluating linear functions a subspace?

Observe:

$$f(x) = ax + b$$

$$f(1) = a \cdot 1 + b$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$f(2) = a \cdot 2 + b$$

$$= \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$f(x_i) = a \cdot x_i + b$$

$$= \begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

So:

$$\begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} f(x_i) \\ \vdots \end{bmatrix} = \begin{bmatrix} x_i & 1 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Fact: Vectors of the form $\begin{bmatrix} f(1) \\ f(2) \\ \vdots \\ f(4) \end{bmatrix}$ <u>are</u> the <u>image</u> of $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$.

So finding the best fit line is equivalent to finding the closest vector in the image of $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$ to

$$\vec{b} = \begin{bmatrix} 0.3 \\ 0.7 \\ 1 \\ 1.3 \end{bmatrix}.$$

This is a problem we know how to do.

☆ Find the Approximate (least squares solution) to $A x = \vec{b}$, $x = \begin{bmatrix} a \\ b \end{bmatrix}$.

Solution to original problem:

want approx. solution to $A x = \vec{b}$.

① Multiply both sides by A^T

$$A^T A x = A^T \vec{b}$$

(consistent).

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.7 \\ 1 \\ 1.3 \end{bmatrix}$$

$$= \begin{bmatrix} 9.9 \\ 3.3 \end{bmatrix}$$

(2) Solve

$$A^T A x = A^T b$$

$$\begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 9.9 \\ 3.3 \end{bmatrix}.$$

Observe $\det(A^T A) = 20 \neq 0$

so $$(A^T A)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix}.$$

$$x = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} \begin{bmatrix} 9.9 \\ 3.3 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 6.6 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= \begin{bmatrix} 0.33 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Unpacking

$$f(x) = ax + b = 0.33x.$$

What's the error

$$\epsilon = \sqrt{\sum_{i=1}^{4} (0.33_i - y_i)^2}$$

$$= \sqrt{(0.03)^2 + \dots}$$

$$= \sqrt{0.003}$$

We did better than our guess.

Key ingredient for using Linear algebra to solve this problem was that the space of vectors of the form

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$$ as f

varies over our class of functions is a subspace.

What classes of functions can we use for curve fitting?

Example: Polynomials of degree $\leq K$.

★Problem: Find a quadratic polynomial $f(x) = ax^2 + bx + c$ which best fits the data points:

x	y
-1	8
0	8
1	4
2	16

.

(Minimizes

$$E := \left(\sum \left(f(x_i) - y_i\right)^2\right)^{1/2}$$

i.e. over all quadratics.)

To use least squares we need

$$\left\{ \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} : f \text{ quadratic} \right\}$$

to be a subspace.

We can show its the <u>image</u> of the following transformation

$$
\begin{bmatrix} \vdots \\ f(x_i) \\ \vdots \end{bmatrix} = \begin{bmatrix} f(-1) \\ f(0) \\ f(1) \\ f(2) \end{bmatrix} = \begin{bmatrix} (-1)^2 & -1 & 1 \\ 0^2 & 0 & 1 \\ 1^2 & 1 & 1 \\ 2^2 & 2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}
$$

$$
f(x) = ax^2 + bx + c
$$

(More generally for f a polynomial of degree $\leq N$

$$
\begin{bmatrix} \vdots \\ f(x_i) \\ \vdots \end{bmatrix} = \begin{bmatrix} x_i^N \cdots x_i^3 & x_i^2 & x_i & 1 \\ & \vdots & & \\ & \vdots & & \end{bmatrix} \begin{bmatrix} c_N \\ \vdots \\ c_1 \\ c_0 \end{bmatrix}.
$$
)

Problem is equivalent to minimizing the distance between

$$\operatorname{Im}\left(\underbrace{\begin{bmatrix} (-1)^2 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}}_{A}\right)$$

and $b = \begin{bmatrix} 8 \\ 8 \\ 4 \\ 16 \end{bmatrix}$.

What is the solution?

① Compute

$$A^T A = \begin{bmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{bmatrix}.$$

$$A^T b = \begin{bmatrix} 76 \\ 29 \\ 36 \end{bmatrix}.$$

② Solve

$$A^T A x = A^T b$$

(by row reduction)

$$\left[\begin{array}{ccc|c} 17 & 8 & 6 & 76 \\ 8 & 6 & 2 & 28 \\ 6 & 2 & 4 & 36 \end{array}\right] \xRightarrow[\text{reduce}]{\text{row}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 5 \end{array}\right].$$

Unpack

$$f(x) = 3x^2 - x + 5$$

Compare exiperimental + predicted values:

x	$f(x)$	y
-1	9	8
0	5	8
1	7	4
2	15	16

$$\epsilon := \sqrt{1^2 + 3^2 + 3^2 + 1^2}$$

$$= \sqrt{20} \sim 4.47.$$

The most general class of functions for which this technique works is the set of functions

$$f = a_1 g_1(x) + a_2 g_2(x) + \dots + a_n g_n(x)$$

for $g_1, \dots, g_n$ some fixed functions. (Span $g_1, \dots, g_n$).

Ex: degree 1 polynomials $g_1(x) = 1$ $g_2(x) = x$.