Linear regression

Topics we'll cover

Cornerstone of Statistics

1. Regression with multiple predictor variables
2. Least-squares regression
3. The least-squares solution

Diabetes study

Data from $n = 442$ diabetes patients.

For each patient:

- 10 features $x = (x_1, \ldots, x_{10})$

  age, sex, body mass index, average blood pressure, and six blood serum measurements.

- A real value y: the progression of the disease a year later.

Regression problem:

- **response** $y \in \mathbb{R}$
- **predictor variables** $x \in \mathbb{R}^{10}$

linear function of $x \to y = wx + b$ $\quad y =$

linear function x single variable $\Rightarrow$ $y = wx + b$

slope, intercept

Least-squares regression

Linear function of 10 variables: for $x \in \mathbb{R}^{10}$,

$$f(x) = w_1x_1 + w_2x_2 + \cdots + w_{10}x_{10} + b = w \cdot x + b$$

where $w = (w_1, w_2, \ldots, w_{10})$. — single vector w

Penalize error using **squared loss** $(y - (w \cdot x + b))^2$.

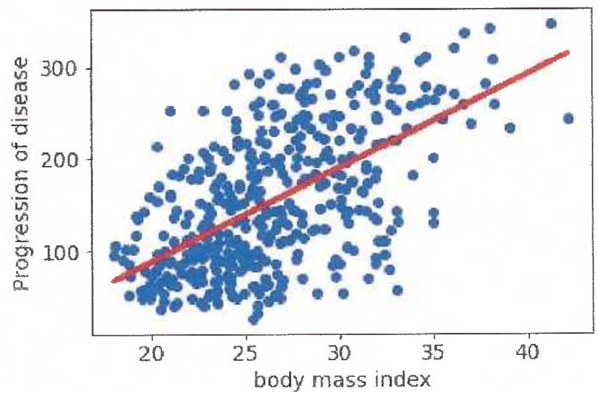
Least-squares regression:

- *Given:* data $(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)}) \in \mathbb{R}^d \times \mathbb{R}$
- *Return:* linear function given by $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$
- *Goal:* minimize the **loss function**

$$L(w, b) = \sum_{i=1}^{n} (y^{(i)} - (w \cdot x^{(i)} + b))^2$$

Back to the diabetes data

- No predictor variables: mean squared error (MSE) = 5930 $\Rightarrow$ variance in y
- One predictor ('bmi'): MSE = 3890



- Two predictors ('bmi', 'serum5'): MSE = 3205
- All ten predictors: MSE = 2860

Least-squares solution 1

Linear function of d variables given by $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$:

$$f(x) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b = w \cdot x + b$$

Assimilate the intercept b into w:

- Add a new feature that is identically 1: let $\widetilde{x} = (1, x) \in \mathbb{R}^{d+1}$ d+1 dimension coz we added 1 to original x

$$\underbrace{(4 \quad 0 \quad 2 \quad \cdots \quad 3)} \Longrightarrow \underbrace{(1 \quad 4 \quad 0 \quad 2 \quad \cdots \quad 3)}$$

- Set $\widetilde{w} = (b, w) \in \mathbb{R}^{d+1}$
- Then $f(x) = w \cdot x + b = \widetilde{w} \cdot \widetilde{x} = (b, w) \cdot (1, x) = b + w \cdot x$

Goal: find $\widetilde{w} \in \mathbb{R}^{d+1}$ that minimizes

$$L(\widetilde{w}) = \sum_{i=1}^{n} (y^{(i)} - \widetilde{w} \cdot \widetilde{x}^{(i)})^2$$

New wts

$\Rightarrow$ rewrite as purely matrix vector product

Least-squares solution 2

$X\tilde{w} \equiv \begin{pmatrix} \tilde{w}\,\bar{x}^{(1)} \\ \tilde{w}\,\bar{x}^{(2)} \\ \vdots \\ \tilde{w}\cdot\bar{x}^{(n)} \end{pmatrix}$

Write

$$X = \begin{pmatrix} \longleftarrow & \widetilde{x}^{(1)} & \longrightarrow \\ \longleftarrow & \widetilde{x}^{(2)} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \widetilde{x}^{(n)} & \longrightarrow \end{pmatrix}, \quad y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix}$$

$n \times (d+1)$ $\quad$ $n \times 1$

$y - X\tilde{w} = \begin{pmatrix} y^{(1)} - \tilde{w}\cdot\hat{x}^{(1)} \\ \vdots \\ y^{(n)} - \hat{w}\,\hat{x}^{(n)} \end{pmatrix}$

the squared error is?

rewrite in term of matrix instead of summation

Then the loss function is

$$\boxed{L(\widetilde{w}) = \sum_{i=1}^{n}(y^{(i)} - \widetilde{w}\cdot\widetilde{x}^{(i)})^2 = \|y - X\widetilde{w}\|^2}$$

and it minimized at $\widetilde{w} = (X^TX)^{-1}(X^Ty)$.

NO SUMMATION

we take derivative to minimize

THIS IS THE BREAD & BUTTER OF STATISTICS

SUMMATION, INSTEAD OF FOR LOOPS,

WE USE A MATRIX

$\tilde{w}$ $\quad$ w tiddle $\rightarrow$ results after assimilating b into w