Monday, 21 October 2024 10:41

Normal Equation

Closed Form Solution to Linear Regression

$Y = \theta_0 + \theta_1 x$

- Y: Target Variable
- θ_0: bias / Intercept
- θ_1: Slope
- x: dependent variable

$J(\theta_0,\theta_1) = \sum^{n} \underbrace{\left(y_i - (\theta_0 + \theta_1 x_i)\right)}_{\text{Residual}}{}^2$

J: Cost function

$n\times 2$

X = Matrix of inputs $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1_n & x_n \end{bmatrix}$

Y = Vector of targets $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}^{n\times 1}$

$\theta = [\theta_0 \ \theta_1] \xrightarrow[\text{Prefer}]{1\times 2} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}^{2\times 1}$

$Y = X^{n\times 2} \theta^{2\times 1}$

$n\times 1 \quad n\times 2 \quad 2\times 1$

$J(\theta) = \frac{1}{n} \| y - X\theta \|^2$

$\|z\|^2 = \sum_i z_i^2 = z^T z$

$J(\theta) = \frac{1}{n} (Y - X\theta)^T (Y - X\theta)$

$= \frac{1}{n}\, \underbrace{Y^T Y}_{\text{sum of squares of targets}} \underbrace{- Y^T X\theta - Y X^T \theta^T}_{\text{cross product real/predicted variables}} + \underbrace{\theta^T X^T X \theta}_{\text{sum of squares of predicted values}}$

$\nabla J(\theta)$ = constant + Linear wrt θ + Quadratic wrt θ

$-Y^T X\theta_j$ → vector of pred. values

$1\times n \cdot n\times 1 \to 1\times 1$ scalar

$\frac{\partial (a^T\theta)}{\partial(\theta)} = a$ → "linear-like" product of two scalars

$\frac{\partial (\theta^T a)}{\partial(\theta)} = a$

$\frac{\partial (-Y^T X\theta)}{\partial(\theta)} = -X^T y$ ($2\times n \quad n\times 1$)

$\frac{\partial (-Y X^T \theta^T)}{\partial(\theta)} = -X^T y$

$\theta^T \underbrace{X^T X}_{} \theta$ ($d\times n \quad n\times d \to d\times d$) → Covariance Matrix of input features. Square Symmetric

Rule : $\theta^T A \theta^T$ → symm. square

$\frac{\partial(\theta^T A \theta)}{\partial(\theta)} = 2A\theta$

$= 2 X^T X \theta$

$\frac{\partial(4x^2)}{\partial x} = 2\cdot 4x$

$X^T X$: sums up how each features in the dataset relates to each other

$J(\theta) = \frac{1}{n} - 2X^T y + 2X^T X\theta$

$= -\frac{2}{n} X^T (Y - X\theta)$

$0 = -\frac{2}{n} X^T (Y - X\theta)$

$X^T (Y - X\theta) = 0$

$X^T y - X^T X\theta = 0$

$X^T y = X^T X \theta$

$\theta = \underbrace{(X^T X)^{-1} X^T}_{\text{pseudo invers}} y$

pseudo invers $X^T (X X^T)^{-1}$

Must be invertible

$X^T X$: ✓ Full rank

Rank of a matrix : # of lineraly independent cols / Rows

x_1 = size in m^2

x_2 = size in $feet^2$

1m = 3,28 feet

$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad Y = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$

Solve it!

A^{-1} for a symm square Matrix $= \frac{1}{\det(A)} \cdot \mathrm{adj}(A)$

$\mathrm{adj}(A) = C^T = \begin{pmatrix} (-1)^{1+1} A_{22} & (-1)^{1+2} A_{21} \\ (-1)^{2+1} A_{12} & (-1)^{2+2} A_{11} \end{pmatrix}^T$

$\det(A) = A_{11} A_{22} - A_{12} A_{21}$ (only for 2x2 Matrices)

→ determinant of the submatrix formed by deleting the ith row and Jth col

also called Minor

$x \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$

$X^T X = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$

$X^T y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 23 \end{bmatrix}$

$(X^T X)^{-1} = \frac{1}{3(14) - 6(6)} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$

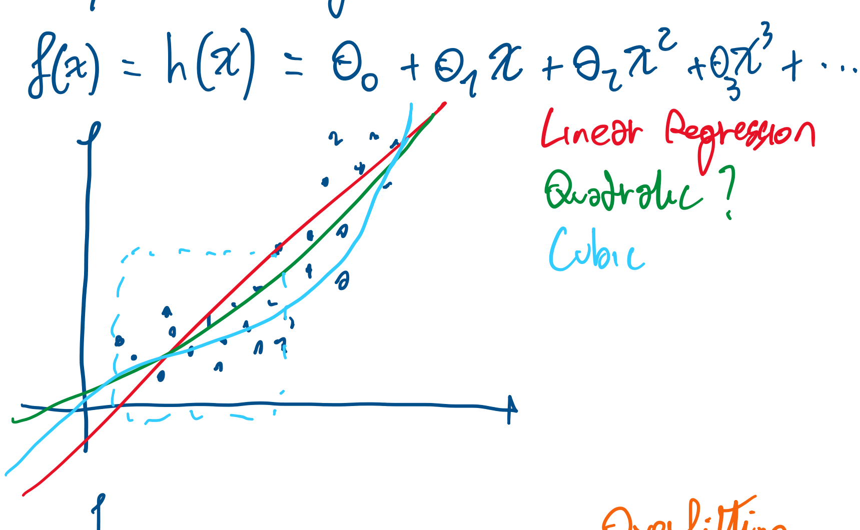
$= \begin{bmatrix} 2.33 & -1 \\ -1 & 0,5 \end{bmatrix}$

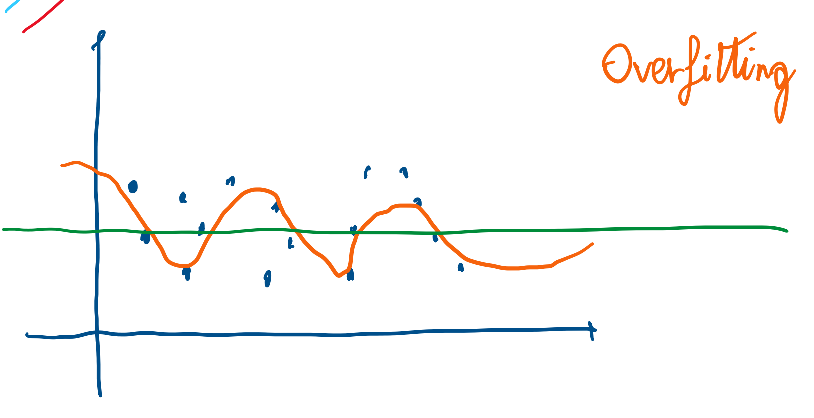
$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 2.33 & -1 \\ -1 & 0.5 \end{bmatrix} \begin{bmatrix} 10 \\ 23 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 1.5 \end{bmatrix}$?

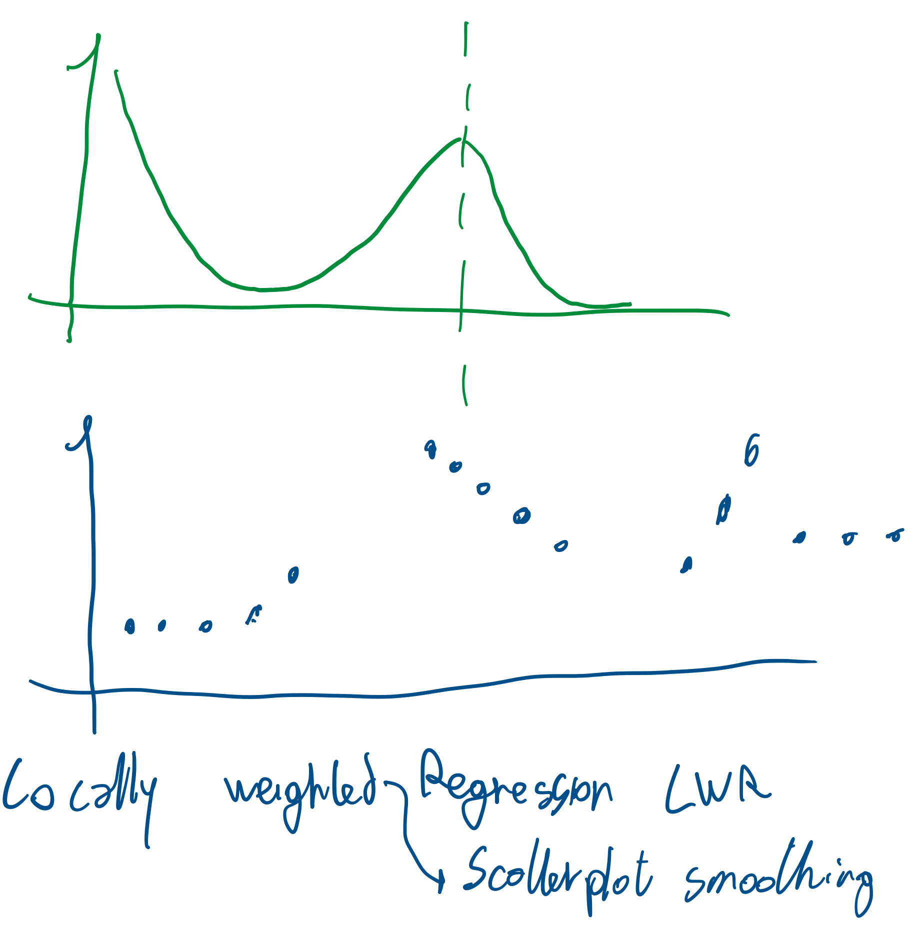
Matrix inversion has almost cubic cost :(

Poly nomial Regression

$f(x) = h(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \dots + \theta_m x^m$







Locally weighted Regression LWR

→ Scatterplot smoothing

Non parametric : no single equation (parameters set) to describe the data

1. Build a local model around query points
2. Focus on nearby data by weighting more points that are closer to query x_q
3. many θ_0, θ_1 as query points

Kernel function : $K(x, x_q)$

$K(x_i, x_q) = w_i(x_q) = \exp\left(-\frac{(x_i - x_q)^2}{2\tau^2}\right)$

τ = band width controls how fast the weight is going to drop with the distance

$J(\theta_0, \theta_1) = \sum^{n} w_i(x_q) \left(f_\theta(x_i) - y\right)^2$

Specific for query point

$x \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad Y \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}$ fit regression at $x = 2.5$ with $\tau = 1$

$x_1 = \exp\left(-\frac{(1-2.5)^2}{2\cdot 1^2}\right) \simeq 0.32$

$x_2 = \exp\left(-\frac{(2-2.5)^2}{2\cdot 1^2}\right) \simeq 0.88$

$x_3 \simeq 0.88$

$x_4 \simeq 0.32$

$\theta^* = (X^T W X)^{-1} (X^T W y)$

$\begin{pmatrix} 0.32 & & & 0 \\ & 0.88 & & \\ & & 0.88 & \\ 0 & & & 0.32 \end{pmatrix}^{4\times 4}$