Agenda

- → Intro. Use Case ✓
- → Data Wrangling
- → Algebraic Intuition
- → Geometric Intuition
- → Optimization
- ✗ Calculation

Today: Predict the sales for a car.

Keywords: Regression

Data. [Kaggle]

$x^{(i)} = [x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \dots x_n^{(i)}]$

Goal:

$y = f(x)$

Tools & Not Tools

Pickup 1000 in last 3 hrs.

Tools: Numpy, Pandas

→ Have to implement

→ Tools are limited.

Sklearn, Tensorflow

→ Automate implementation

Terminologies

"car" → Inputs → "Price."

Sale $\rightarrow y$

$\hat{y} = f(x)$

$\hat{y} = (w_1 x + w_0)$ ← $y = f + \text{Noise}$

Dependent variable (Output) — Independent variable (Input) — Weight — Intercept

$w_1 = 0$

$w_0 = 8$

$\boxed{\hat{y} = 0x + 8}$

$\boxed{y = 8}$

Assuming no noise.

Stationary → 2 features i.e. "age", "odometer"

$w = \text{Noise}$ for in 3D!

→ Error.

$\hat{y} = w_1 x_1 + w_2 x_2 + w_0$

$\log_e y = y_1 + y_2 + \log_e$

$w_1 = -5$ $\quad w_2 = 3$ $\quad w_0 = 7$

$\boxed{\hat{y} = -5x_1 + 3x_2 + 7}$

$y = mx + c$

$y = w_1 x + w_0$

$\boxed{w_1 = 5}$

$\boxed{w_0 = 7}$

$y = mx + c$

$y = w_1 x + w_0$

$\boxed{w_1 = 4}$

$\boxed{w_0 = 7}$

Sklearn → 4 features

$\hat{y} = f$ hypothesis.

$\boxed{\hat{y} = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_0}$

x_1 x_2 x_3 x_4

$\hat{y} = f(x)$ is linear.

$\hat{y} = f(x)$ is linear.

$\boxed{(d+1)\text{ dim}}$

1 input feature → 2D space fitter

2 input features → 3D space fitter.

d input features → $(d+1)$ dim space fitter.

$f(x) \rightarrow \langle w_0, w_1, w_2 \dots w_n \rangle$

"Weights"

Using optimization

$\pi^{(d+1)}$ "hyperplane"

1 dimension

$\hat{y} = w_1 x + w_0$

d input features → $(d+1)$ weights $w_0 \dots w_n$

$\boxed{\hat{y} = w_1 x + w_0}$

Loss function.

Goal: $\mathcal{D} = \{(x^{(1)}, y^{(1)}), \dots\}$: $x^{(i)} \in \mathbb{R}^d$: $y^{(i)} \in \mathbb{R}$

Find: $\hat{y} = f(x)$

Possible → $w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_0$

$\rightarrow [w_1, w_2 \dots w_n]^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + w_0$

$\boxed{\hat{y} = w^T x + w_0}$

actual — model error

$[1, 2, 3, 4]^T$

Find such w's $y = (w^T x + w_0)$

$\boxed{y^{(i)} \approx \hat{y}^{(i)}} \quad \forall i = 1 \dots n$

actual value — predicted value

Find w's minimize difference the actual & predicted price

$\boxed{y^{(i)} - \hat{y}^{(i)}} = e^{(i)}$

actual — $f(x^{(i)})$ — +ve — −ve

SL $-5.1k \Rightarrow (-10k)$

SL $-0.65k \Rightarrow (+5k)$

$\underset{w, w_0}{\operatorname{argmin}} \left[\sum_{i=1}^{n} e^{(i)} \right]$ Find such w & w_0 that minimum the sum of errors.

Error with: $\underset{w, w_0}{\operatorname{argmin}} \sum_{i=1}^{n} e^{(i)}$

$e^{(1)} \rightarrow$ +ve

$e^{(2)} \rightarrow$ −ve

$\Rightarrow e^{(1)} + e^{(2)} \dots e^{(n)}$

$E \Rightarrow 0$ ✗

Sol: Take Square. $[e^{(i)}]^2$

$\underset{w, w_0}{\operatorname{argmin}} \sum_{i=1}^{n} [e^{(i)}]^2$

$$\boxed{L = \underset{w, w_0}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{n}}_{\text{Sum}} \underbrace{[y^{(i)} - (w^T x^{(i)} + w_0)]^2}_{\text{Square}}}$$

MSE → (Mean Square Error)

Optimization.

$L(w, w_0) = \sum_{i=1}^{n} (y^{(i)} - \hat{y}^{(i)})^2$ $\quad \hat{y}^{(i)} = [w_0 x^{(i)} + w_0]$

Sum — Square

$\frac{\partial L}{\partial w_0} \Rightarrow \frac{\partial}{\partial w_0} (y - w_1 x_1 - w_0)^2$

$\frac{\partial L}{\partial w_0} \Rightarrow \frac{\partial}{\partial w_0} (y - \hat{y})^2$

$2 \cdot (y - w_1 x_1 - w_0) \cdot \frac{d(y - w_1 x_1 - w_0)}{dw_0}$

$2 \cdot (y - w_1 x_1 - w_0) \cdot (-1)$

$2 \cdot (y - \hat{y}) \cdot (-1)$

$2 \cdot (y - y)$

$\boxed{\frac{\partial L}{\partial w_0} \Rightarrow \frac{2}{n} \sum_{i=1}^{n} (\hat{y} - y)}$

$z = (y - \hat{y})$

$\frac{\partial z^2}{\partial z} \cdot \frac{\partial z}{\partial w_0}$

$\frac{\partial z}{\partial w_0} \cdot \frac{\partial (y - w_0 x_1)}{\partial w_0}$

$\frac{\partial z}{\partial w_0} \cdot (-x_1)$

$2 \cdot (y - \hat{y}) \cdot (-x_1)$

$2 \cdot (\hat{y} - y) \cdot x_1$

$\boxed{\frac{\partial L}{\partial w_0} = \frac{2}{n} \sum_{i=1}^{n} (\hat{y} - y) \cdot x_1}$

Gradient Descent.

① randomly init(s)

② Repeat n-times times (100)

$w_j^{old} = w_j^{old} - \eta \cdot \boxed{\frac{\partial L}{\partial w_j}}_{(w_j^{old})}$

$w_0^{new} = w_0^{old} - \eta \cdot \boxed{\frac{\partial L}{\partial w_0}}_{(w_0^{old})}$

$w_1^{new} = w_1^{old} - \eta \cdot \boxed{\frac{\partial L}{\partial w_1}}_{(w_1^{old})}$

Derivative

$(-x - 5)^2$

$f(x_1), f(x_2), f(x_3) \dots f(x_n)$ ✗

$y = f(x)$ ✓

Weight

Height

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \text{ "Delta"}$$