Linear Regression

Recap: supervised learning technique used to predict continuous numerical values by learning input variables (/features) & an output variable (/target).

Ex1. ⇒ Determining CGPA [0–10] from factors like hours of sleep, coffee consumption before exam, avg. hours of class attention etc.

Ex2. ⇒ Determining house price (in ₹) from factors like house area (in m^2).

Bias

Let's Start!

x_i = i^{th} feat. value

$$\hat{y} = \underline{\underline{\theta_0}} + \underline{\underline{\theta_1}} x_1 + \underline{\underline{\theta_2}} x_2 + \dots \theta_n x_n$$

→ θ_j = parameter

→ linear eq! Just like $y = mx + c$.

→ $y = w_1 x_1 + \dots w_n x_n + b$

→ Predicted value [E.M.] # features

Life is all 'bout vectors & matrices!

Note: x_i are independent var!

$$\hat{y} = h_\theta(x) = \vec{\theta} \cdot \vec{x} \quad [E.M.]$$

↳ Hypothesis fxn.
(proposed explanation for a phenomenon)

$$= \theta^T x$$

Dummy Example: Predict Insurance charges (… ?)

Given: 1) Age 2) Sex 3) BMI
4) Children 5) Smoker
6) Region (… ?)

✓ **Step 1:** Get your data!

✓ **Step 2:** Divide your data

Train (~80%) Test (~20%)

[give an analogy.]

Step 3 : Find the model params. (i.e. θ_j)

But, I have no idea...

Errors ⟵ i) Take random values,

ii) & optimize over iterations!

How?

Step 4 : cost fxn. (to minimize errors & optimize θ)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y)^2$$

why?

* parabola.

$$\boxed{\text{Obj : Find } \theta_0, \theta_j \text{ s.t. } \hat{y}_i \approx y_i \;\; \forall (x_i, y_i)}$$

Hint: I want to minimize $J(\theta)$

↓

Step 5 : Gradient Descent

$$\boxed{\frac{\partial J(\theta_0, \theta_i)}{\partial \theta}}$$

$$\left.\begin{aligned}
\theta_0 &= \theta_0 - \alpha \frac{\partial J}{\partial \theta_i} \\
\text{(Age)} \quad \theta_1' &= \theta_1 - \alpha \frac{\partial J}{\partial \theta_i} \\
\text{(BMI)} \quad \theta_2' &= \theta_2 - \alpha \frac{\partial J}{\partial \theta_i}
\end{aligned}\right\} \text{MULTI-VARIATE L.R.}$$

α = learning rate = step size
(scaling factor)

Regulates how much of network's parameters are updated at every iteration of OPTIMIZATION.

Then...

$$\left.\begin{aligned}
\theta_1 &= \theta_1' \\
\theta_2 &= \theta_2'
\end{aligned}\right\} \text{update.}$$

* let's do it ...

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$\frac{\partial J}{\partial \theta_1} = \underbrace{2 \cdot \sum_{i=1}^{m} (\underline{\underline{\hat{y}_i}} - y_i) \cdot x_{1i}}$$

$$\theta_1 = \theta_1 - \frac{1}{2m} \cdot$$

$$\Rightarrow \theta_1 = \theta_1 - \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) \cdot x_{1i}$$

$$\Rightarrow \theta_2 = \theta_2 - \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) \cdot x_{2i}$$

$$\Rightarrow \theta_0 = \theta_0 - \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i) \cdot 1$$