08/08/24

MAP

MLE

(Maximum likelihood estimate).

n data points.

$D = \{x_1, x_2, \dots, x_n\}$

likelihood of these data points, given a model $\theta = \{ \quad \}$

$$\log \underline{\hat{P}(D|\theta)} = \hat{P}(\{x_1, x_2, \dots x_n\}|\theta) = \boxed{\prod_{k=1}^{n} \hat{P}(x_k|\theta)}$$

the MLE of θ is : $\theta_{MLE} = \arg\max_{\theta} \hat{P}(D|\theta)$

Coin flip

H | T

$$\boxed{\theta = P(H) = q = \frac{\# \text{ of } H}{\text{total count}} = \frac{n_1}{n_1 + n_2}}$$

$D = \{x_1, x_2, \dots x_n\}$

$n_1 : H$

$n_2 : T$

$$P(D|\theta) = \prod_{i=1}^{n} P(x_i|\theta) = q^{n_1}(1-q)^{n_2}$$

To obtain $\hat{q} = \arg\max_{q} q^{n_1}(1-q)^{n_2} = F$

$$\frac{\partial F}{\partial q} = n_1 q^{n_1 - 1}(1-q)^{n_2} - q^{n_1}(1-q)^{n_2 - 1} \times n_2 = 0$$

$$\Rightarrow q^{n_1-1}(1-q)^{n_2-1}(n_1(1-q) - q n_2) = 0$$

$$\Rightarrow q = \frac{n_1}{n_1 + n_2}$$

$$\log \hat{P}(D|\theta) = \sum_{i=1}^{N} \log(x_i|\theta) \qquad D = \{x_1 \dots x_n\}$$

MAP : Maximum a posteriori Estimate

$$\log(\theta_{MAP}) = \arg\max_{\theta} P(\theta|D) = \arg\max_{\theta} \underbrace{P(D|\theta)}_{\text{likelihood}} \underbrace{P(\theta)}_{\text{prior}}$$

Posterior prob

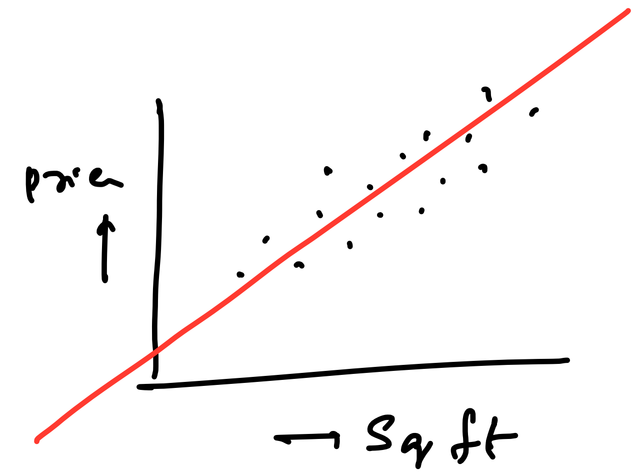
$$= \log P(\theta) + \log P(D|\theta)$$

$$= \log P(\theta) + \sum_{i=1}^{n} P(x_i|\theta)$$

Linear Regression

Living area (Sqft)	price ($)
2104	400
1600	330
3000	232
→ 2500	?

$|f(x)-y|$



$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots$

$= \theta_0 + \sum_{i=1}^{m} \theta_i x_i$

x_0

$= \sum_{i=0}^{m} \theta_i x_i$

$x_0 = 1$

minimize the distance from all the points to the line L

$f(x_1 x_2 x_3) = x_1 + 3x_2 + 4x_3$

$(1, 2, 3)$

$f = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 =$

Loss: $\frac{1}{2}\sum_{i=1}^{m} (f(x_i) - y_i)^2$

$\theta = \{\theta_0, \theta_1, \dots, \theta_m\}$

$d_i = (x_i, y_i)$

d_1
d_2
d_3
$\vdots$
d_m

$J(\theta) = \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2$

m = # of training instance.

Superscript → instance

j = Subscript → = parameter / feature

$\frac{\partial J}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{2} (h_\theta(x^i) - y^i)^2$

$= 2 \times \frac{1}{2} (h_\theta(x^i) - y^i) \frac{\partial}{\partial \theta_j} (h_\theta(x) - y)$

$= (h_\theta(x^i) - y^i) \times \frac{\partial}{\partial \theta_j} (\sum_{k=1}^{n} \theta_k x_k - y)$

$= (h_\theta(x^i) - y^i) x_j$

$\hat{\theta}_j \leftarrow \theta_j - \eta \frac{\partial J}{\partial \theta_j}$

new, old, learning rate

$\frac{\partial J}{\partial \theta_p} = (h_\theta(x^i) - y^i) x_p$

Batch GD | Stochastic GD | mini Batch GD

$$\hat{\theta}_j \leftarrow \theta_j - \eta \frac{\partial J}{\partial \theta_j}$$

$\theta = 5$