COL 774
Machine Learning
Aug 18, 2021

[Sunday Aug 22
8am - 9am]

Last class:-

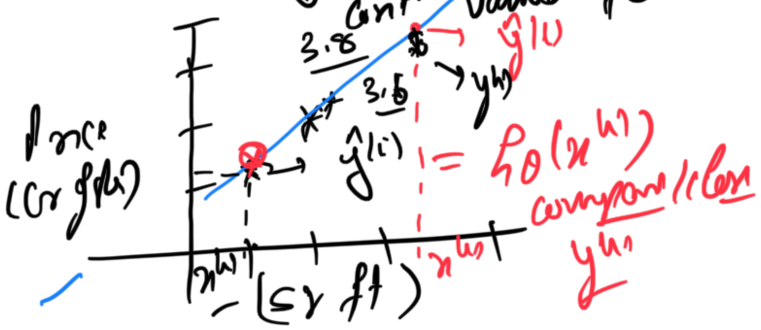
Semi-supervised
Reinforcement learning.

→ Linear Regression :- $h_\theta(x)$ cont.

$x \in R^n$ | $y \in R$

cont valued features

[overfitting | underfitting]



$x^{(i)}$	$y^{(i)}$
1093	
2086	

Data actually is generated by a linear fn

$$\left[y = \theta_1^* x_1 + \theta_0^* + (\varepsilon) \right] \rightarrow \text{noise (or unmodeled effects)}$$

slope ⇓ Intercept

$$h_\theta(x) = \theta_1 x_1 + \theta_0 \;]$$

↑ hypothesis fn

$h_\theta \in \mathcal{H}$
⇓
All linear functions (lines) in 2-D

# of parameters that need to be estimated = 2

HD More general:-

$\mathcal{H}$:- linear

$$h_\theta(x) = \underbrace{\sum_{j=1}^{n} \theta_j x_j + \theta_0}_{n} \Rightarrow$$

corresponds to equation of a hyperplane

$$= \sum_{j=0}^{n} \theta_j x_j \qquad \boxed{x_0 = 1} \to \text{intercept}$$

① → What is our hypothesis space?] linear

② → What is a good hypothesis?

which "fits" (data) well

↳ How do you characterize this
→ intuitively;
→ mathematically

↳ When do we say that a line "fits" data well?

↳ Loss Fn = $\underbrace{\frac{1}{m}\sum_{i=1}^{m} \overbrace{|y^{(i)} - \hat{y}^{(i)}|}^{\Leftarrow \text{ not differentiable}}}_{\Downarrow}$ $\hat{y}^{(i)} = h_\theta(x^{(i)})$

Find that θ which "minimizes" ⟹ this

closeness of the fit

⟹ Instead: if we squared error; → that has a more natural justification under some modeling assumptions ($x^{(i)}$'s are random, $y^{(i)}$'s)

Loss(θ)

Loss fn = $\boxed{-L_2(\theta)}$

error fn

$$J(\theta) = \underbrace{1/m \sum_{i=1}^{m} [y^{(i)} - \hat{y}^{(i)}]^2}_{\text{MSE}}$$

$$= 1/m \sum^{m} [y^{(i)} - h_\theta(x^{(i)})]^2$$

$\frac{1}{m} \sum_{i=1} [y^{(i)} - h_\theta(x^{(i)})]$

$\Downarrow$

"kernel trick"

② what is a good fit

$\Rightarrow \underline{J(\theta)} = \frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} - \sum_{j=0}^{n} \theta_j x_j \right]^2 \cdot \frac{d}{d\theta}$

③ How do we find the "optimal" hypothesis

$\underset{\theta}{\operatorname{argmin}} \; \underline{J(\theta)}$

$\left[\underset{\theta \in \mathbb{R}^{n+1}}{\operatorname{argmin}} \; J(\theta) \right]$

$\underset{\theta \in \mathbb{R}^{n+1}}{\min} \; J(\theta)$

$\underset{\theta}{\operatorname{argmin}} \; \frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} - h_\theta(x^{(i)}) \right]^2$

$(\theta_n \ldots \theta_0)$

MSE

$(x_n \ldots x_1)$

$h_\theta(x) = \sum_{j=1}^{n} \theta_j x_j + \theta_0$

$\equiv \sum_{j=0}^{n} \theta_j x_j \quad (x_0 = 1)$

$\Rightarrow \theta = \begin{bmatrix} \theta_n \\ \vdots \\ \theta_0 \end{bmatrix} \Rightarrow$ set of parameters in the linear model

$\sum^{n} \theta_i x_i = [\theta^T x]$

$$x = \begin{bmatrix} x_n \\ | \\ | \\ x_0 = 1 \end{bmatrix} \qquad \sum_{j=0} \qquad = x^T\theta$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - h_\theta(x^{(i)})\right)^2 \qquad \sum_{j=0}^{n}\theta_j x_j^{(i)}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[\underbrace{y^{(i)} - \theta^T x^{(i)}}\right]^2 = \text{quadratic fn (in } \theta)$$

$\Rightarrow$ as a fn., find the argmin of $J(\theta)$

$\Rightarrow$ Optimizing this quadratic fn. (of $(n+1)$ parameters)

↳ Detour:- finding minima of function

(univariate) multivariate 2-variate fn

$$f(\theta) = (\theta-3)^2 + 1 \quad [\text{local minima} / \text{maxima}] \qquad f(\theta) = \theta_1^2 + \theta_2^2 + 2\theta_1\theta_2 + 5\theta_1 + 6$$

$\Rightarrow$ $\theta = 3$

$f'(\theta)$

$$\left[\frac{\partial f(\theta)}{\partial \theta}\right] = 0$$

gradient

$\Rightarrow$ $\underset{\theta}{\operatorname{argmin}} f(\theta)$ $\qquad \Rightarrow \underset{\theta}{\operatorname{argmin}} f(\theta)$

↳ θ [finding analytically]

② Gradient Descent