COL774
Machine Learning
Aug 24, 2021

Last class:- Linear Regression:

$h_\theta(x)$:- Linear Function.

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

$\theta^T x^{(i)}$

Gradient update Rule → optimization

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta \cdot \nabla_\theta f(\theta) \rightarrow \text{gradient}$$

$\in \mathbb{R}^{n+1}$

if $\theta \in \mathbb{R}^{n+1}$

$$f(\theta) = \theta_1^2 + \theta_2^2 + 3\theta_1\theta_2 + 2\theta_2 + 3$$

$\nabla f(\theta)$ 2 eq.

Loss fn:- $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left[\left[y^{(i)} - h_\theta(x^{(i)}) \right]^2 \right]$ Analytical ?

\# $\arg\min_\theta J(\theta)$:- ⟹ can be done by using gradient descent.

Gradient Descent: 1

$$\nabla_\theta J(\theta) = \nabla_\theta \left(\frac{1}{2m} \sum_{i=1}^{m} \left[y^{(i)} - \theta^T x^{(i)} \right]^2 \right]$$

$$= \frac{1}{2m} \sum^{m} \nabla_\theta \left[y^{(i)} - \theta^T x^{(i)} \right]^2$$

$$= \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \cancel{2}\left[y^{(i)} - h_\theta(x^{(i)})\right] \nabla_\theta \left(y^{(i)} - \theta^T x^{(i)}\right)$$

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

$$\nabla_\theta J(\theta)$$

$$\left[x^{(i)}\right] \in \mathbb{R}^{n+1}$$

As der

$$\sum_{j=0}^{n} \theta_j x_j = \theta^T x^{(i)}$$

$$\frac{\partial}{\partial \theta_j} \sum_{j=0}^{n} \theta_j x_j = x_j$$

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

$$\theta^T x^{(i)}$$

$$\nabla_\theta \theta^T x^{(i)}$$

$$x^{(i)} \quad \left[\frac{\partial\left[\theta^T x^{(i)}\right]}{\partial \theta_j} \right] \quad \begin{array}{c} x_0 \\ x_j \\ x_n \end{array}$$

Gradient Descent for learning parameters of Linear Regression

$\Rightarrow$ Parameter update Rule

do{ $t=0$; $\theta^{(t)} \leftarrow$ init();

$$\underset{\in \mathbb{R}^{n+1}}{\theta^{(t+1)}} \leftarrow \theta^{(t)} - \eta \cdot \nabla_\theta J(\theta)$$

$$\leftarrow \theta^{(t)} + \eta \cdot \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right) x^{(i)}$$

$$= \underset{\in \mathbb{R}^{n+1}}{\theta^{(t)}} + \eta \cdot \frac{1}{m} \underbrace{\sum_{i=1}^{m} x^{(i)}\left(y^{(i)} - \theta^T x^{(i)}\right)}_{\mathbb{R}^{n+1}}$$

$t \leftarrow t+1$;

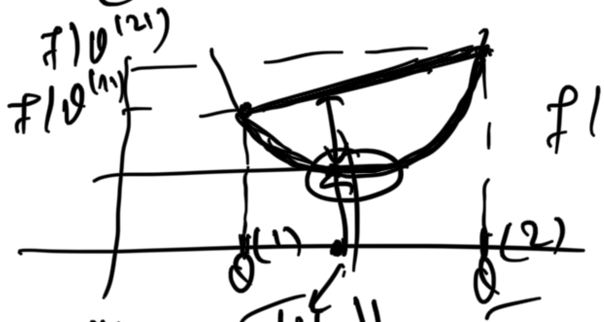
} while (converged)

Convexity :- (Convex function) | SGD : Nothing more efficient GD

$f(\theta)$:-

$f : R \to R$ | $f : R^m \to R$

Convex function



$\alpha\theta^{(1)} + (1-\alpha)\theta^{(2)}$ ⇓

$f(\theta) = (\theta-3)^2 + 1$:- strictly convex

↳ $\frac{\partial^2 f(\theta)}{\partial^2 \theta} \geq 0$ ✓

θ, $\theta^{(1)}, \theta^{(2)} \in R$

f is convex iff :- $(0 \leq \alpha \leq 1)$

$$\left[f(\underbrace{\alpha\theta^{(1)} + (1-\alpha)\theta^{(2)}}_{\text{Convex combo}}) \leq \alpha f(\theta^{(1)}) + (1-\alpha) f(\theta^{(2)}) \right.$$

↓ convex

strictly convex $0 < \alpha < 1$

Generalize to $f : R^{(m+1)} \to R$

f is convex iff

$$\left[f(\alpha\theta^{(1)} + (1-\alpha)\theta^{(2)}) \leq \alpha f(\theta^{(1)}) + (1-\alpha) f(\theta^{(2)}) \right.$$

$\theta^{(1)}, \theta^{(2)} \in R^{m+1}$

Convex set :-

C :- $\theta^{(1)}, \theta^{(2)}$

$\theta^{(1)}, \theta^{(2)} \in C$

$\alpha\,\theta^{(1)} + (1-\alpha)\,\theta^{(2)} \in C$

Last Bit:- ~~How~~ Understanding convexity in terms of second order derivatives.

$f: R \rightarrow R$

$$\frac{\partial^2 f(\theta)}{\partial^2 \theta} \geq 0$$

$f: R^n \rightarrow R$

$\Downarrow$

$\left[\frac{\partial f(\theta)}{\partial \theta_j}\right]$ $\nabla_\theta f(\theta)$ 1st derivative

$\xrightarrow{n \times 1}$ H Hessian matrix 2nd derivative

$n \times 1$

$$\frac{\partial^2 f(\theta)}{\partial \theta_j \partial \theta_k}$$

H_{jk}

$$\equiv \frac{\partial^2 f(\theta)}{\partial \theta_k \partial \theta_j}$$

Convexity & Hessian matrix

If $f: R^n \rightarrow R$ is convex then the corresponding Hessian matrix H is +ve semi-definite

B:- symmetric

→ A matrix $B \in R^{(n+1)\times(n+1)}$ is +ve semi-definite

if ...

$$\forall z \in \mathbb{R}^{n+1} \quad \underbrace{z^T B z}_{\text{Quadratic form}} \geq 0$$

$$\underbrace{\sum_{j,k} z_j B_{jk} z_k}_{} = z^T B z$$

$$\overset{z^T}{[\quad]}\;\overset{B}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}\;\overset{z}{\begin{bmatrix} \\ \\ \\ \end{bmatrix}} \geq 0$$

If $f: \mathbb{R}^{n+1} \to \mathbb{R}$ is convex & doubly differentiable then Hessian H must be +ve semi-definite

$$z^T H z \geq 0 \quad \forall z \in \mathbb{R}^{n+1}$$

Homework exercise -]

↳ We will be able to show that

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left[y^{(i)} - \theta^T x^{(i)} \right]^2$$

is a convex function in θ parameters.

$\nabla_0 J(0)$ I

H Hessian matrix $\in R^{(n+1)\times(n+1)}$

$z^T H z \geq 0 \quad \forall z \in R^{(n+1)}$

Concave ✓