

DEEP LEARNING

by Ian Goodfellow.

Jasua Bengio.

Aaron Courville.

$$J(\theta) = \frac{1}{q} \sum [f^*(x) - f(x, \theta)]^2$$

$$f(x; \theta) = f(x; w, b) = x^T w + b$$

$$J(\theta) = \frac{1}{q} \sum_{x \in X} [f^*(x) - (x^T w + b)]^2$$

$$\frac{\partial J(\theta)}{\partial w} = -\frac{1}{q} 2(x^T w + b) \cdot x^T = -\frac{1}{q} x^T (x^T w + b)$$

$$\frac{\partial J(\theta)}{\partial b} = -\frac{1}{q} 2(x^T w + b) \cdot 1 = -\frac{1}{q} (x^T w + b)$$

MINIMISE MSE

$$MSE_{test} = \frac{1}{m} \sum_i (\hat{y}^{(test)} - y^{(test)})_i^2, \text{ with } m = \text{number example}$$

We can see that $MSE_{test} = 0$ when $\hat{y}^{(test)} = y^{(test)}$

We can see also that

$$MSE_{test} = \frac{1}{m} \|\hat{y}^{(test)} - y^{(test)}\|_2^2$$

Let Minimize $MSE_{train} \Rightarrow$ solve where $\nabla_w MSE_{train} = 0$

$$\|\hat{y}\|_2 = \|y\| = x^T x$$

$$\nabla_w MSE_{train} = 0$$

$$\Rightarrow \nabla_w \left[\frac{1}{m} \|\hat{y}^{(train)} - y^{(train)}\|_2^2 \right] = 0$$

$$\nabla_w \|\hat{y}^{(train)} - y^{(train)}\|_2^2 = 0$$

$$\nabla_w \|X^{(train)} w - y^{(train)}\|_2^2 = 0 \quad \text{with } \hat{y}^{(train)} = X^{(train)} w$$

$$\nabla_w (X^{(train)} w - y^{(train)})^T (X^{(train)} w - y^{(train)}) = 0$$

$$\begin{aligned} \nabla_w & \left[W^T X^{(train)T} X^{(train)} W - W^T X^{(train)T} y^{(train)} - y^{(train)T} X^{(train)} W + y^{(train)T} y^{(train)} \right] = 0 \\ & - 2W^T X^{(train)T} y^{(train)} \end{aligned}$$

$$\nabla_w (w^T X^{(\text{train})T} X^{(\text{train})} w - 2 w^T X^{(\text{train})T} y^{(\text{train})} + y^{(\text{train})T} y^{(\text{train})}) = 0$$

$$2 X^{(\text{train})T} X^{(\text{train})} w - 2 X^{(\text{train})T} y^{(\text{train})} = 0$$

$$\Rightarrow w = (X^{(\text{train})T} X^{(\text{train})})^{-1} X^{(\text{train})T} y^{(\text{train})}$$

known as normal equations

$$\# J(\theta) = \frac{1}{4} \sum_{x \in X} (f^*(x) - f(x; \theta))^2$$

$$f(x; \theta) = f(x; w, b) = x^T w + b$$

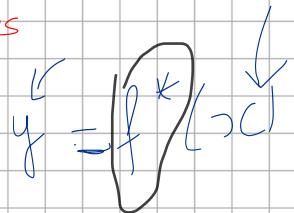
$$\nabla \left[\frac{1}{4} \sum (f^*(x) - (x^T w + b))^2 \right] = 0$$

$$\Rightarrow \nabla \|f^*(x) - (x^T w + b)\|_2^2 = 0$$

$$\Rightarrow \nabla [f^*(x) - (x^T w + b)]^T [f^*(x) - (x^T w + b)] = 0$$

$$\Rightarrow \nabla [f^*(x)^T f^*(x) - f^*(x)(x^T w + b) - (x^T w + b)^T f^*(x) + (x^T w + b)^T (x^T w + b)] = 0$$

X	Y	XOR
0	0	0
1	0	1
0	1	1
1	1	0



suppose: // $f^{(1)}(x) = w^T x$
 $f^{(2)}(h) = h^T w$

$$h = f^{(1)}(x; w, c) \text{ and } y = f^{(2)}(h; w, b)$$

output of the network

$$\Rightarrow f(x) = f^{(2)}(f^{(1)}(x; w, c)) = f^{(2)}(w^T x) = (w^T x)^T w = \boxed{x^T w w} = (x^T w w)^T = \boxed{w^T w^T x}$$

$$= x^T w' \text{ where } w' = w w$$

putting an activation function "g" = ReLU $\Rightarrow g(z) = \max\{0, z\}$

Considering the bias term $f(b)$ can be written as;

$$f(x; w^1, b^1, w^2, b^2) = f^{(2)}(f^{(1)}(x; w^1, b^1)) = f^{(2)}(f^{(1)}((w^1)^T x + b^1) + b^2)$$

$$= f^{(2)}(\max\{0, (w^1)^T x + b^1\} + b^2)$$

$$= g((w^2)^T \max\{0, (w^1)^T x + b^1\} + b^2)$$

$$= \max\{0, \max\{(w^2)^T \max\{0, (w^1)^T x + b^1\} + b^2\}\}$$

