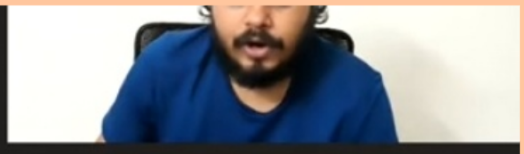
Gradient Descent

Monday, June 21, 2021 1:53 PM



n input

Rows $= m$ Columns $= n$

	1	2	3		n	y	$\hat{y}$
1	x_{11}	x_{12}	x_{13}	.....	x_{1n}	y_1	$\hat{y}_1$
2	x_{21}	x_{22}	x_{23}	.....	x_{2n}	y_2	
⋮	⋮						
m	x_{m1}	x_{m2}	x_{m3}	.....	x_{mn}	y_m	

$$\{w_1 \quad w_2 \quad w_3 \ldots w_n \quad w_0\}$$

$(n+1)$

$$\sigma(w_1 x_{11} + w_2 x_{12} + w_3 x_{13} + \ldots + w_n x_{1n} + w_0) = \hat{y}_1$$

$$\sigma(w_1 x_{21} + w_2 x_{22} + w_3 x_{23} + \ldots + w_n x_{2n} + w_0)$$

$\hat{y}$

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix} = \begin{bmatrix} \sigma(w_0 + w_1 x_{11} + w_2 x_{12} + \ldots + w_n x_{1n}) \\ \sigma(w_0 + w_1 x_{21} + w_2 x_{22} + \ldots + w_n x_{2n}) \\ \vdots \\ \sigma(w_0 + w_1 x_{m1} + w_2 x_{m2} + \ldots + w_n x_{mn}) \end{bmatrix}$$

$$\hat{y} = \sigma\left(\begin{bmatrix} w_0 + w_1 x_{11} + w_2 x_{12} + \ldots + w_n x_{1n} \\ w_0 + w_1 x_{21} + w_2 x_{22} + \ldots + w_n x_{2n} \\ \vdots \\ w_0 + w_1 x_{m1} + w_2 x_{m2} + \ldots + w_n x_{mn} \end{bmatrix}\right)$$

$$\hat{y} = \sigma\left(\overset{A}{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & & \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}} \overset{B}{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix}}\right)$$

$$\boxed{\hat{y} = \sigma(\vec{X}\,\vec{W})}$$

→ Now we need to convert our loss function into matrix from.

$$L = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$$

$$\Rightarrow L = -\frac{1}{m}\left[\sum y_i \log(\hat{y}_i) + \sum (1-y_i)\log(1-\hat{y}_i)\right]$$

$$\sum y_i \log(\hat{y}_i) = y_1 \log \hat{y}_1 + y_2 \log \hat{y}_2 \cdots y_m \log \hat{y}_m$$

$$= \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_m \end{bmatrix} \begin{bmatrix} \log \hat{y}_1 \\ \log \hat{y}_2 \\ \vdots \\ \log \hat{y}_m \end{bmatrix}$$

$$= \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_m \end{bmatrix} \log\left(\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{bmatrix} \right)$$

$$= \vec{y} \log \vec{\hat{y}_i}$$

$$\boxed{= \vec{y} \log\left(\sigma(\vec{x}\vec{m})\right)}$$

For,

$$\sum (1-y_i) \log(1-\hat{y}_i)$$

$$= \boxed{(1-\vec{y}) \log\left(\sigma(1-\vec{x}\vec{w})\right)}$$

Finally we have,

$$L = -\frac{1}{m}\left[\vec{y} \log \vec{\hat{y}} + (1-\vec{y}) \log(1-\vec{\hat{y}})\right]$$

where $\hat{y} = \sigma(\vec{x}\vec{M})$

→ Now will perform gradient descent on it.

$$W = W - \alpha \frac{\Delta L}{\Delta W}$$

Now,

$$\frac{dL}{dW} = ?$$

For, $\vec{y} \log \hat{\vec{y}}$

$$\frac{d}{dw} y \log \hat{y} = y \frac{d}{dw} \log \hat{y}$$

$$= \frac{y}{\hat{y}} \frac{\partial}{\partial L} (\hat{y})$$

$$= \frac{y}{\hat{y}} \frac{d}{dL} \underline{\sigma(wx)}$$

$$= \frac{y}{\hat{y}} \underline{\sigma(wx)[1 - \sigma(wx)]} \frac{d}{dw}(wx)$$

$$= \frac{y}{\hat{y}} \hat{y}(1 - \hat{y}) x$$

$$= y(1-\hat{y})x$$

For

$$\frac{d}{dw}(1-y)\log(1-y)$$

$$= (1-y)\frac{d}{dw}\log(1-\hat{y})$$

$$= \frac{1-y}{1-\hat{y}}\frac{d}{dw}(1-\hat{y})$$

$$= -\frac{1-y}{1-\hat{y}}\frac{d}{dw}\sigma(wx)$$

$$= -\frac{(1-y)}{(1-\hat{y})}\left(\sigma(wx)(1-\sigma(wx))\right)\frac{d}{dw}(wx)$$

$$= -\frac{1-y}{\cancel{1-\hat{y}}}\hat{y}\cancel{(1-\hat{y})}x$$

$$= -\hat{y}(1-y)x$$

Together,

$$\frac{d}{dw} = -\frac{1}{m}\left[y(1-\hat{y})x - \hat{y}(1-y)x\right]$$

$$= -\frac{1}{m}\left[y(1-\hat{y}) - \hat{y}(1-y)\right]x$$

$$= -\frac{1}{m}\left(y - \cancel{y\hat{y}} - \hat{y} + \cancel{y\hat{y}}\right)x$$

$$= -\frac{1}{m}\left(y - \hat{y}\right)x$$

so,

$$\vec{w} = \vec{w}_0 + \alpha \frac{1}{m}(y - \hat{y})x$$