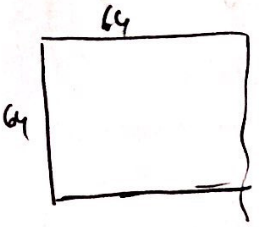
19.01.2019

Coursera Machine Learning Working Notes

Binary Classification

$x \rightarrow y$

$$x = \begin{bmatrix} 255 \\ 231 \\ \vdots \\ \\ \end{bmatrix}$$



$64 \times 64 \times 3 = 12288$

$n = n_x = 12288$

Notation

$(x, y) \quad x \in \mathbb{R}^{n_x}, \; y \in \{0, 1\}$

m training examples : $(x^{(1)}, y^{(1)}, \ldots \ldots (x^{(m)}, y^{(m)})$

$m = m_{train}$

m_{test} = number of test ex.

$$X = \begin{bmatrix} | & | & | \\ x^{(1)} & x^{(2)} & x^{(m)} \\ | & | & | \end{bmatrix} n_x$$

$\longleftrightarrow m$

X.shape = (n_x, m)

$$Y = [y^{(1)}, y^{(2)} \ldots, y^{(m)}] \quad y \in \mathbb{R}^{1 \times m}$$

Y.shape = $(1, m)$

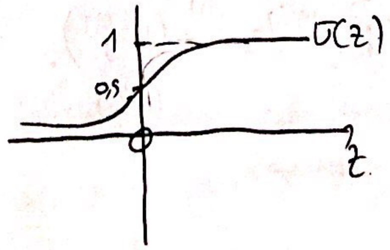
logistic Regression

Given x, want $\hat{y} = P(y=1|x)$ → olma ihtimali

Parameters : $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$

Output : $\hat{y} = w^T x + b$ => lineer regresyon için uygun ama binary clas. için 0 veya 1 olmalı.

Output binary clas : $\hat{y} = \sigma(w^T x + b)$ $\quad 0 \leq \hat{y} \leq 1$ => binary için böyle olmalı



$\sigma(z) = \frac{1}{1+e^{-z}}$

w ve B yi ayarlamaya çalışıyoruz

If z large $\sigma(z) \approx \frac{1}{1+0} = 1$

If z small, negative number $\sigma(z) = \frac{1}{1+e^{-z}} \approx \frac{1}{1+\text{Bignum}} = 0$

...nallikle W ve B parametresi ayrı tutulur.

$x_0 = 1$, $x \in \mathbb{R}^{n_x+1}$

$\hat{y} = \sigma(\theta^T x)$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_x} \end{bmatrix} \begin{array}{l} \} b \leftarrow \\ \\ \Big\} w \leftarrow \\ \\ \\ \end{array}$$

başka bir notasyon

Logistic Regression Cost Function

- W ve B parametrelerini eğitmek için, cost function tanımlanmalı.

$\hat{y} = \sigma(w^T x + b)$, where $\sigma(z) = \dfrac{1}{1+e^{-z}}$

Given $\{(x^{(1)}, y^{(1)}), \ldots (x^{(m)}, y^{(m)})\}$ want $\hat{y}^{(i)} \approx y^{(i)}$

üstte i varsa i'inci distirme örneği notasyonu.

Loss function:

$\mathcal{L}(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$ => gradient decent not work well

$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1-y) \log(1-\hat{y}))$ => logistic regression için

If $y = 1$: $\mathcal{L}(\hat{y}, y) = -\log \hat{y}$ $\hat{y} \rightarrow$ mümkün olduğunca büyük olmalı

$\log \hat{y} \rightarrow$ büyük olmalı.

$\rightarrow$ loss function

if $y = 0$ $\mathcal{L}(\hat{y}, y) = -\log(1-\hat{y})$ $\log(1-\hat{y})$ büyük olmalı

$\hat{y}$ küçük olmalı

Cost Function : $J(w, b) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \cdot \log(1-\hat{y}^{(i)})$

loss function computes the error for a single training example,

the const function is the average of the loss function of the entire training set.

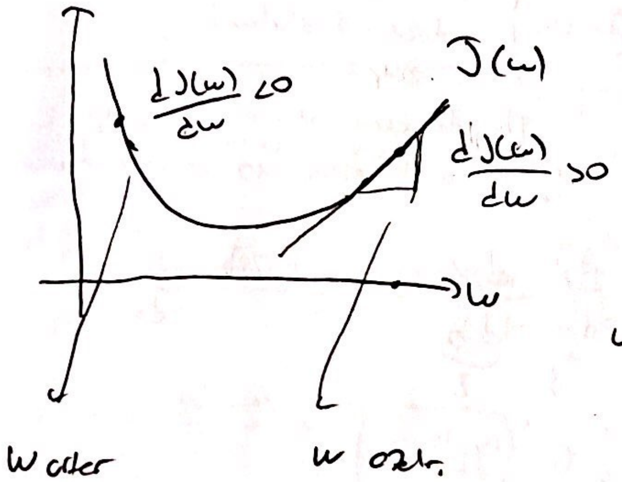
Gradient Descent

19.04.22

Recap: $\hat{y} = \sigma(w^T x + b)$, $\sigma(z) = \frac{1}{1+e^{-z}}$

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\log \hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

Want to find w,b that minimize J(w,b)



b'yi ignore ettik

Repeat {

$w := w - \alpha \frac{dJ(w)}{dw}$ — learning rate

update w

$\frac{dJ(w)}{dw}$ / dw

}

Converge'ye yakın tekrar edecek.

$w := w - \alpha\, dw$

$J(w,b)$ $\quad w := w - \alpha \frac{dJ(w,b)}{dw}$ $\quad \partial \rightarrow$ partial derivative depends on one parameter slope

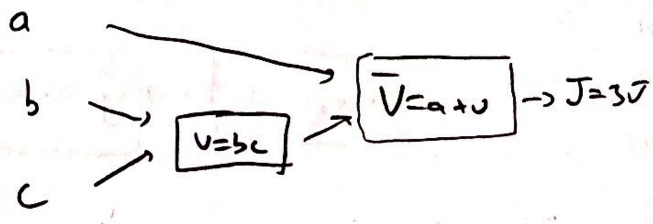
$b := b - \alpha \frac{dJ(w,b)}{db}$ $\quad y = \log_a u(x) \Rightarrow y' = \frac{u'(x)}{u(x)} \cdot \log_a e$

$J(a,b,c) = 3(a+bc)$

$u = bc$

$V = a+u$

$J = 3V$



One step of backward propagation on a computation graph yields derivative of final output variable

Computing derivatives

"da" = 3

"dv" => Çıktı değerin türevi

$a=5$

$b=3$

$c=2$

$\boxed{u=bc} \xrightarrow{6} \boxed{v=a+u} \xrightarrow{11} \boxed{J=3v} \; 33$

← türev alma

backward propagation

$$\frac{dJ}{dv}=3$$

$$\frac{dJ}{da}=3$$

$a=5 \to 5.001$

$v=11 \to 11.001$

$J=33 \to 33.003$

$$\frac{dJ}{da}=\underbrace{\frac{dJ}{dv}}_{3}\underbrace{\frac{dv}{da}}_{1}$$

Chain Rule

$$\underbrace{\frac{dJ}{dv}}_{3}\underbrace{\frac{dv}{du}}_{1} \quad \text{"du"}=3$$

$\Rightarrow \boxed{\frac{\text{Final Out Var}}{d\,var}} \Rightarrow$ dvar (Kodlarken) output

The derivative of a final output variable with respect to various intermediate quantities

$$\frac{dJ}{db}=\underbrace{\frac{dJ}{du}}_{3}\underbrace{\frac{du}{db}}_{2} \qquad u=2b \quad \frac{du}{db}=2$$

$$\frac{dJ}{db}=6 \quad \text{"db"}=6$$

$$\frac{dJ}{dc}=\underbrace{\frac{dJ}{du}}_{3}\underbrace{\frac{du}{dc}}_{3} \qquad u=3c \quad \frac{du}{dc}=3$$

$$\frac{dJ}{dc}=9 \quad \text{"dc"}=9$$

Logistic Regression recap

$z=w^Tx+b$

$\hat{y}=a=\sigma(z)$

$\mathcal{L}(a,y)=-(y\log(a)+(1-y)\log(1-a))$

x_1, w_1, x_2, w_2, b → $\boxed{z=w_1x_1+w_2x_2+b} \to \boxed{\hat{y}=a=\sigma(z)} \to \boxed{\mathcal{L}(a,y)}$

19.01.2

Logistic Regression derivatives

x_1, w_1, x_2, w_2, b $\rightarrow$ $\boxed{z = w_1x_1 + w_2x_2 + b}$ $\rightarrow$ $\boxed{a = \sigma(z)}$ $\rightarrow$ $\boxed{\mathcal{L}(a,y)}$

$dz = \frac{d\mathcal{L}}{dz} = \frac{d\mathcal{L}(a,y)}{dz}$

$dz = (a-y)$

$\frac{d\mathcal{L}(a,y)}{da}$

"da"

$da = \frac{-y}{a} + \frac{1-y}{1-a}$

$\mathcal{L}(a,y) = -(y \log a + (1-y)\log(1-a))$

$\frac{d\mathcal{L}(a,y)}{da} = -\left(y \cdot \frac{1}{a} + (1-y)\frac{-1}{(1-a)}\right)$

$= \frac{-y}{a} + \frac{(1-y)}{(1-a)}$

loss derivative respect to a

$\boxed{\frac{d\mathcal{L}(a,y)}{da} = \frac{-y}{a} + \frac{1-y}{1-a}}$

$\frac{da}{dz} = a(1-a)$ => Kontrol et

$dz = \frac{d\mathcal{L}}{da} \cdot \frac{da}{dz} = (a-y)$

$a = \frac{1}{1+e^{-z}}$

$\frac{d\mathcal{L}}{dz} = \underbrace{\frac{d\mathcal{L}}{da}}_{\frac{-y}{a}+\frac{1-y}{1-a}} \cdot \underbrace{\frac{da}{dz}}_{a(1-a)}$

$\frac{d\mathcal{L}}{dz} = a-y$

$\frac{d\mathcal{L}}{dw_1}$ = "dw_1" $= x_1 \cdot dz$

"dw_2" $= x_2 dz$

$db = dz$

$w_1 := w_1 - \alpha dw_1$

$w_2 := w_2 - \alpha dw_2$

$b := b - \alpha db$

$\frac{\partial \mathcal{L}}{\partial w_1}$ = "dw_1" $= x_1 dz$

"dw_2" $= x_2 dz$

"db" $= dz$

$w_1 := w_1 - \alpha dw_1$

$w_2 := w_2 - \alpha dw_2$

$b := b - \alpha db$

Scanned with CamScanner

...stic Regression on m examples

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}(a^{(i)}, y^{(i)})$$ $(x^{(i)}, y^{(i)})$

$$a^{(i)} = \hat{y}^{(i)} = \sigma(z^{(i)}) = \sigma(w^T x^{(i)} + b)$$ $dw_1^{(i)}, dw_2^{(i)}, db^{(i)}$

$$\frac{\partial}{\partial w_1} J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \underbrace{\frac{\partial}{\partial w_1}\mathcal{L}(a^{(i)}, y^{(i)})}_{dw_1^{(i)} - (x^{(i)}, y^{(i)})}$$

$J=0, \quad dw_1=0, \quad dw_2=0, \quad db=0$

For $i=1$ to m

$z^{(i)} = w^T x^{(i)} + b$

$a^{(i)} = \sigma(z^{(i)})$

$J_{+} = -[y^{(i)} \log a^{(i)} + (1-y^{(i)}) \log(1-a^{(i)})]$

$dz^{(i)} = a^{(i)} - y^{(i)}$

$dw_1 \mathrel{+}= x_1^{(i)} dz^{(i)}$

$dw_2 \mathrel{+}= x_2^{(i)} dz^{(i)}$ $n=2$

$db \mathrel{+}= dz^{(i)}$

$J /= m$

$dw_1 /= m; \quad dw_2 /= m; \quad db /= m;$

$$dw_1 = \frac{\partial J}{\partial w_1}$$

$w_1 := w_1 - \alpha dw_1$

$w_2 := w_2 - \alpha dw_2$

$b := b - \alpha db$

Scanned with CamScanner

19.01.20

Backward Propa. Proof.

$\hat{y} = a$

$x_1, w_1, x_2, w_2, b \longrightarrow \boxed{z = w_1 x_1 + w_2 x_2 + b} \longrightarrow \boxed{a = \sigma(z)} \longrightarrow \boxed{\mathcal{L}(a,y)}$

$$\mathcal{L}(a,y) = -\left(y \log a + (1-y)\log(1-a)\right)$$

$$\frac{\mathcal{L}(a,y)}{da} = -\left(y\frac{1}{a} + (1-y)\frac{-1}{(1-a)}\right)$$

$$\boxed{\frac{\mathcal{L}(a,y)}{da} = \frac{-y}{a} + \frac{(1-y)}{(1-a)}}$$

$$\boxed{\text{"}da\text{"} = \frac{-y}{a} + \frac{(1-y)}{1-a}}$$

$a = \frac{1}{1+e^{-z}}$

$\frac{da}{dz} = (1+e^{-z})^{-1} \rightarrow -(1+e^{-z})^{-2} \cdot -e^{-z}$

$\underbrace{(1+e^{-z})^{-2} e^{-z}}$

$\underbrace{a^2 \cdot \left(\frac{1}{a}-1\right)}$

$\underbrace{a - a^2}$

$a(1-a)$

$a = (1+e^{-z})^{-1}$

$a^2 = (1+e^{-z})^{-2}$

$a^{-1} = 1+e^{-z}$

$a^{-1} - 1 = e^{-z}$

$\frac{1}{a} - 1$

$$\boxed{\frac{da}{dz} = a(1-a)}$$

$$\frac{\mathcal{L}(a,y)}{dz} = \underbrace{\frac{d\mathcal{L}(a,y)}{da}} \cdot \frac{da}{dz} = \frac{d\mathcal{L}(a,y)}{dz} = \text{"}dz\text{"}$$

$$\underbrace{\left[\frac{-y}{a} + \frac{(1-y)}{1-a}\right] \cdot \left[a(1-a)\right]}_{a-y}$$

$$\frac{\mathcal{L}(a,y)}{dz} = a - y$$

$$\boxed{\text{"}dz\text{"} = a - y}$$

w_1 turn

$\frac{dz}{dw_1} = x_1$

$$\frac{\mathcal{L}(a,y)}{dw_1} = \underbrace{\frac{d\mathcal{L}(a,y)}{da} \cdot \frac{da}{dz}}_{a-y} \cdot \underbrace{\frac{dz}{dw_1}}_{x_1}$$

d

"dw_1"

$$\frac{\mathcal{L}(a,y)}{dw_1} = \overset{dz}{\overbrace{(a-y)}}\, x_1$$

$\frac{}{dw_1}$

$dw_1 = x_1 dz$

$dw_2 = x_2 dz$

$db = dz$

$$\boxed{\begin{aligned} w_1 &:= w_1 - \alpha\, dw_1 \\ w_2 &:= w_2 - \alpha\, dw_2 \\ b &:= b - \alpha\, db \end{aligned}}$$