CSE 4705 - Artificial Intelligence

Classification - Logistic Regression

Joe Johnson M.S., Ph.D., A.S.A.

Department of Computer Science and Engineering
University of Connecticut

UCONN | UNIVERSITY OF CONNECTICUT

Table of Contents

Table of Contents

Table of Contents

Classification - Logistic Regression - Motivation

Classification

Question	Answer "y"
Is this email spam?	
Is the transaction fraudulent?	
Is the tumor malignant?	

Classification - Logistic Regression - Motivation

Classification

Question	Answer "y"
Is this email spam?	no yes
Is the transaction fraudulent?	
Is the tumor malignant?	

Classification - Logistic Regression - Motivation

Classification

Question	Answer "y"	
Is this email spam?	no	yes
Is the transaction fraudulent?	no	yes
Is the tumor malignant?	no	yes

Classification - Logistic Regression - Motivation

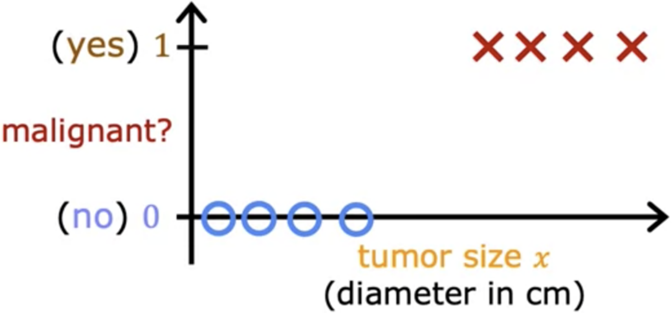
Classification

Question	Answer "y"	
Is this email spam?	no	yes
Is the transaction fraudulent?	no	yes
Is the tumor malignant?	no	yes

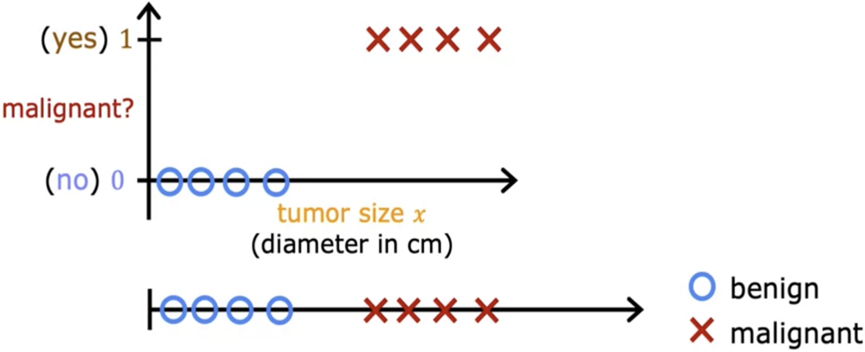
y can only be one of two values

"binary classification"

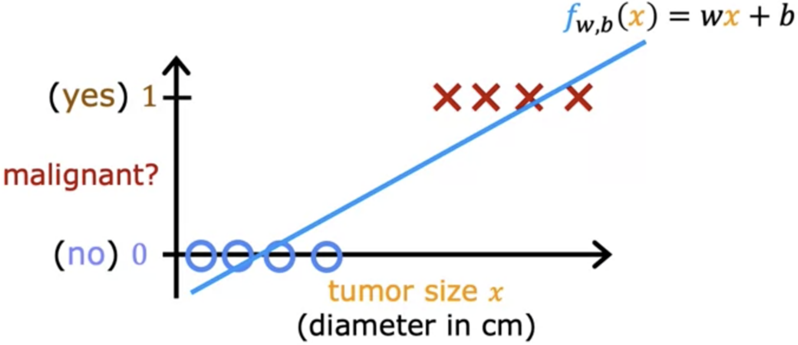
Classification - Logistic Regression - Motivation



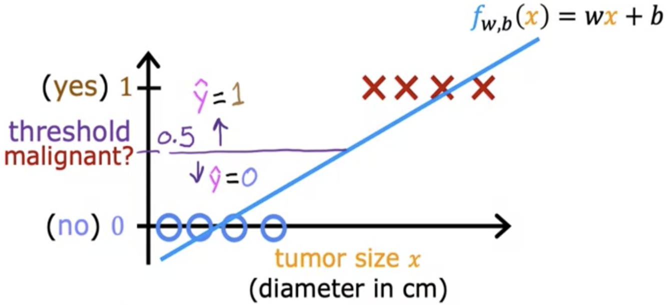
Classification - Logistic Regression - Motivation



Classification - Logistic Regression - Motivation



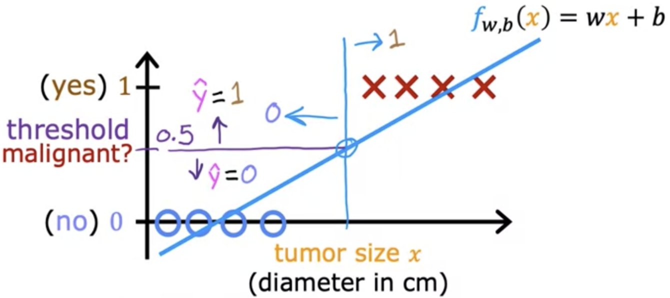
Classification - Logistic Regression - Motivation



$$\text{if } f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$$

$$\text{if } f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$$

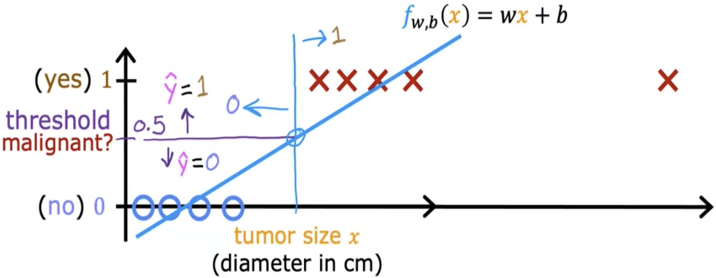
Classification - Logistic Regression - Motivation



if $f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$

if $f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$

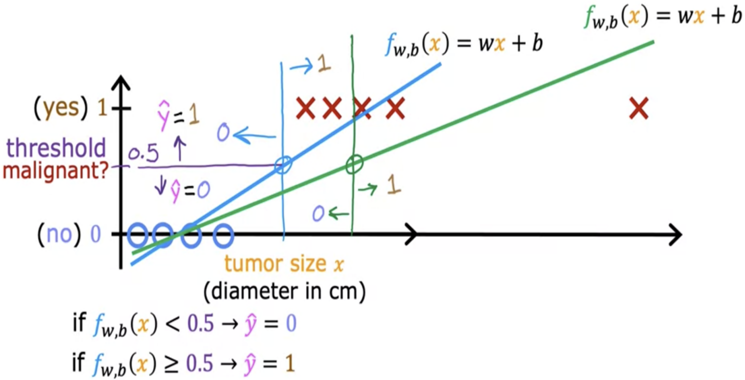
Classification - Logistic Regression - Motivation



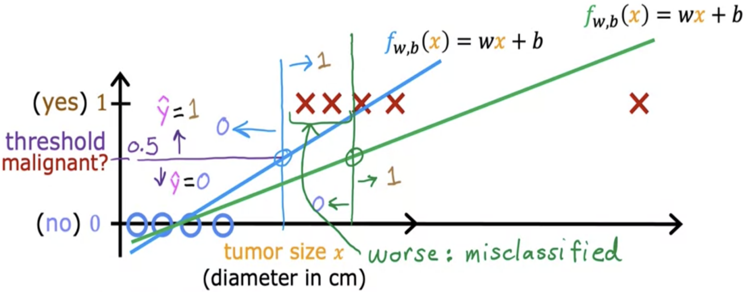
if $f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$

if $f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$

Classification - Logistic Regression - Motivation



Classification - Logistic Regression - Motivation

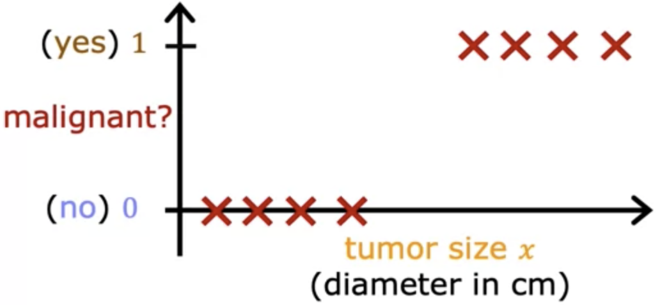


if $f_{w,b}(x) < 0.5 \rightarrow \hat{y} = 0$

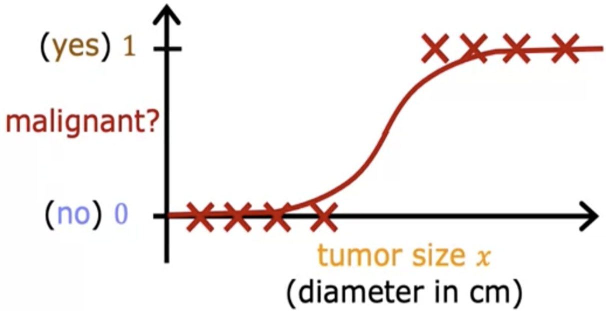
if $f_{w,b}(x) \geq 0.5 \rightarrow \hat{y} = 1$

Table of Contents

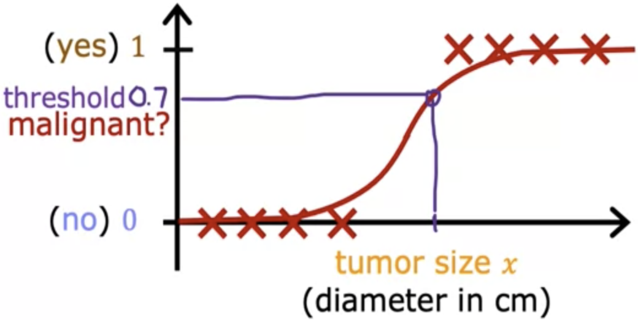
Logistic Regression - Model



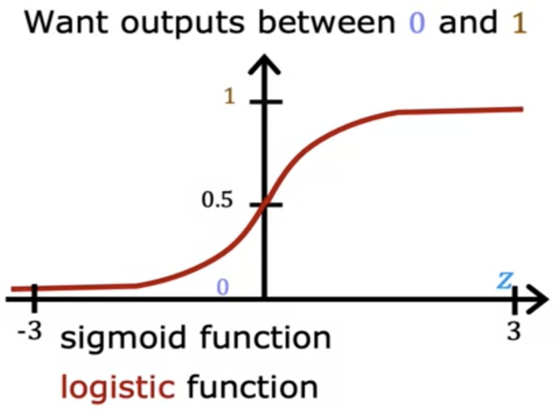
Logistic Regression - Model



Logistic Regression - Model



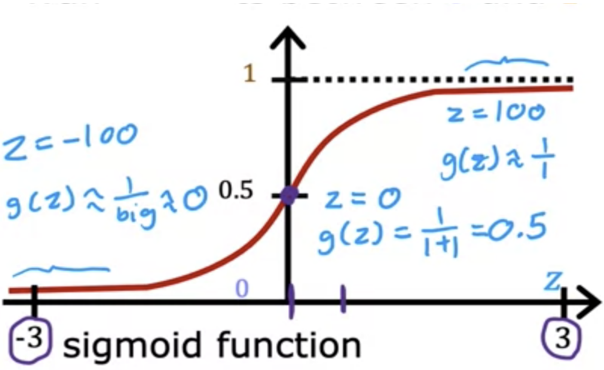
Logistic Regression - Model



Logistic Regression - Model

$$g(z) = \frac{1}{1+e^{-z}} \qquad 0 < g(z) < 1$$

Logistic Regression - Model



Logistic Regression - Model

$$f_{\vec{w},b}(\vec{x})$$

$$\boxed{z = \vec{w} \cdot \vec{x} + b}$$

$$\downarrow$$

$$z$$

$$\downarrow$$

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

Logistic Regression - Model

Interpretation of logistic regression output

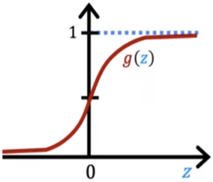
$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

"probability" that class is 1

Example:
x is "tumor size"
y is 0 (not malignant)
or 1 (malignant)

$f_{\vec{w},b}(\vec{x}) = 0.7$
70% chance that y is 1

Logistic Regression - Introduction



$$f_{\vec{w},b}(\vec{x})$$

$$\boxed{z = \vec{w}\cdot\vec{x} + b}$$

$$\downarrow$$

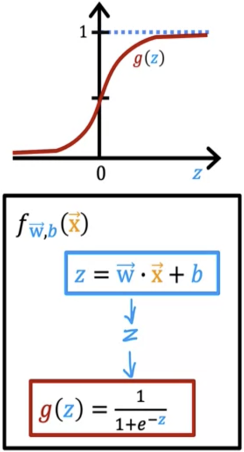
$$z$$

$$\downarrow$$

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

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w}\cdot\vec{x} + b}_{z}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

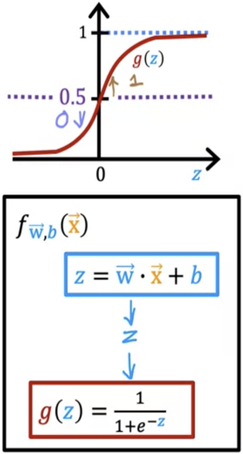
Logistic Regression - Introduction



$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x}+b)}}$$

$$= P(y = 1 | x; \vec{w}, b)$$

Logistic Regression - Introduction



$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

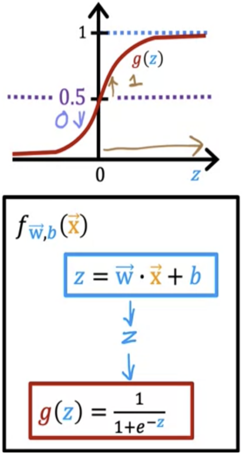
$$= P(y = 1 | x; \vec{w}, b) \quad 0.7 \quad 0.3$$

0 or 1? threshold

Is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

Yes: $\hat{y} = 1$ No: $\hat{y} = 0$

Logistic Regression - Introduction



$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w} \cdot \vec{x} + b}_{z}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$$

$$= P(y = 1 | x; \vec{w}, b) \quad 0.7 \quad 0.3$$

0 or 1? threshold

Is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

Yes: $\hat{y} = 1$ No: $\hat{y} = 0$

When is $f_{\vec{w},b}(\vec{x}) \geq 0.5$?

$$g(z) \geq 0.5$$

$$z \geq 0$$

$$\vec{w} \cdot \vec{x} + b \geq 0$$

Logistic Regression - Introduction

$$f_{\vec{w},b}(\vec{x}) = g(\underbrace{\vec{w}\cdot\vec{x}+b}_{z}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

$$= P(y=1|x;\vec{w},b) \quad 0.7 \quad 0.3$$

0 or 1? threshold

$$\text{Is } f_{\vec{w},b}(\vec{x}) \geq 0.5?$$

Yes: $\hat{y}=1$ No: $\hat{y}=0$

$$\text{When is } f_{\vec{w},b}(\vec{x}) \geq 0.5?$$

$$g(z) \geq 0.5$$

$$z \geq 0$$

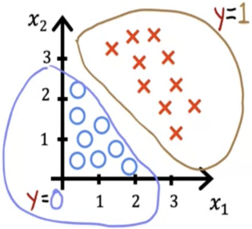
$$\vec{w}\cdot\vec{x}+b \geq 0 \qquad \vec{w}\cdot\vec{x}+b < 0$$

$$\hat{y}=1 \qquad \hat{y}=0$$

Table of Contents

Decision Boundary

Decision boundary

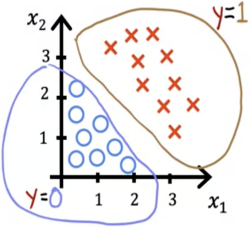


Decision Boundary

Decision boundary

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\underset{1}{w_1}x_1 + \underset{1}{w_2}x_2 + \underset{-3}{b})$$

Decision boundary $z = \vec{w} \cdot \vec{x} + b = 0$



Decision Boundary

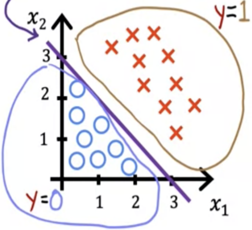
Decision boundary

$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\overbrace{\underset{1}{w_1}x_1 + \underset{1}{w_2}x_2 + \underset{-3}{b}})$$

Decision boundary $z = \vec{w}\cdot\vec{x} + b = 0$

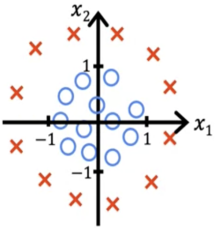
$z = x_1 + x_2 - 3 = 0$

$x_1 + x_2 = 3$



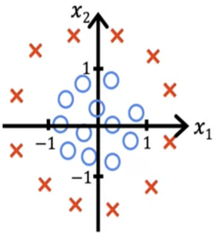
Decision Boundary

Non-linear decision boundaries



Decision Boundary

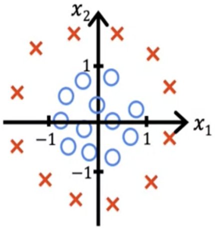
Non-linear decision boundaries



$$\overbrace{w_1 x_1^2 + w_2 x_2^2 + b}^{z}$$

Decision Boundary

Non-linear decision boundaries

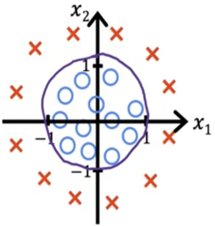


$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\overbrace{\underset{1}{w_1}x_1^2 + \underset{1}{w_2}x_2^2 + \underset{-1}{b}}^{z})$$

decision boundary $z = x_1^2 + x_2^2 - 1 = 0$

Decision Boundary

Non-linear decision boundaries



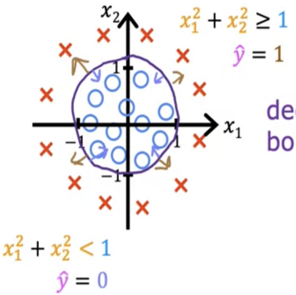
$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\overbrace{\underset{1}{w_1}x_1^2 + \underset{1}{w_2}x_2^2 + \underset{-1}{b}}^{z})$$

decision boundary $z = x_1^2 + x_2^2 - 1 = 0$

$$x_1^2 + x_2^2 = 1$$

Decision Boundary

Non-linear decision boundaries



$$f_{\vec{w},b}(\vec{x}) = g(z) = g(\overbrace{\underset{1}{w_1} x_1^2 + \underset{1}{w_2} x_2^2 + \underset{-1}{b}}^{z})$$

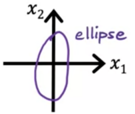
decision boundary

$$z = x_1^2 + x_2^2 - 1 = 0$$

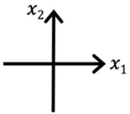
$$x_1^2 + x_2^2 = 1$$

Decision Boundary

Non-linear decision boundaries

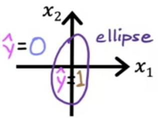


$$f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_1x_2 + w_5x_2^2 + w_6x_1^3 + \cdots + b)$$



Decision Boundary

Non-linear decision boundaries



$$f_{\vec{w},b}(\vec{x}) = g(z) = g(w_1x_1 + w_2x_2 \\ +w_3x_1^2 + w_4x_1x_2 + w_5x_2^2 \\ +w_6x_1^3 + \cdots + b)$$

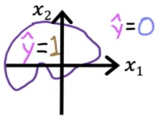


Table of Contents

Table of Contents

Cost Function Motivation

Training set

tumor size (cm)	...	patient's age	malignant?
10		52	1
2		73	0
5		55	0
12		49	1
...		...	...

Cost Function Motivation

Training set

	tumor size (cm) x_1	...	patient's age x_n	malignant? y
$i=1$	10		52	1
⋮	2		73	0
	5		55	0
	12		49	1
$i=m$	...		...	...

$i = 1, \ldots, m$ ← training examples

$j = 1, \ldots, n$ ← features

$\boxed{\text{target } y \text{ is } 0 \text{ or } 1}$

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w}\cdot\vec{x}+b)}}$$

How to choose $\vec{w} = [w_1 \quad w_2 \quad \cdots \quad w_n]$ and b?

Cost Function Motivation

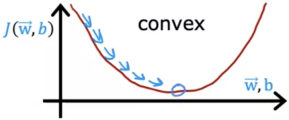
Squared error cost

$$J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m} \boxed{\frac{1}{2}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)^2}$$

loss $\boxed{L\left(f_{\vec{w},b}(\vec{x}^{(i)}), y^{(i)}\right)}$

linear regression

$$f_{\vec{w},b}(\vec{x}) = \vec{w}\cdot\vec{x} + b$$



logistic regression

$$f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$$

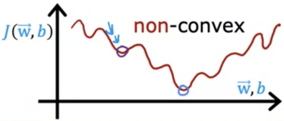


Table of Contents

Form of Cost Function for Logistic Regression

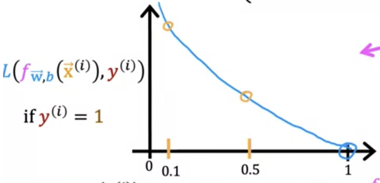
Logistic loss function

$$L\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

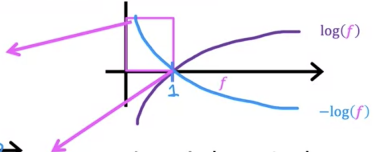
Form of Cost Function for Logistic Regression

Logistic loss function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$



$f_{\vec{w},b}(\vec{x}^{(i)})$



As $f_{\vec{w},b}(\vec{x}^{(i)}) \rightarrow 1$ then loss $\rightarrow 0$

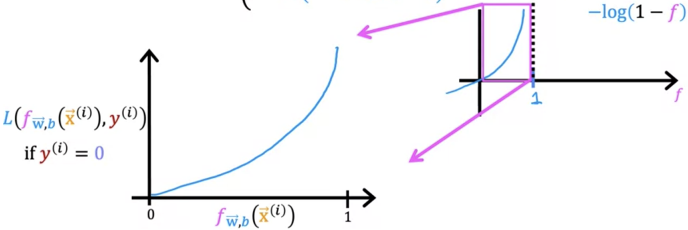
As $f_{\vec{w},b}(\vec{x}^{(i)}) \rightarrow 0$ then loss $\rightarrow \infty$

Loss is lowest when $f_{\vec{w},b}(\vec{x}^{(i)})$ predicts close to true label $y^{(i)}$.

Form of Cost Function for Logistic Regression

Logistic loss function

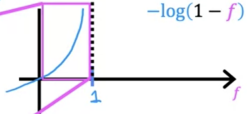
$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$



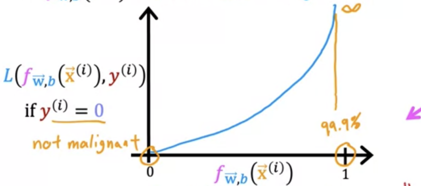
Form of Cost Function for Logistic Regression

Logistic loss function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$



As $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 0$ then loss $\to 0$



As $f_{\vec{w},b}\left(\vec{x}^{(i)}\right) \to 1$ then loss $\to \infty$

The further prediction $f_{\vec{w},b}\left(\vec{x}^{(i)}\right)$ is from target $y^{(i)}$, the higher the loss.

Form of Cost Function for Logistic Regression

Cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m} L\underbrace{\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)}_{\text{loss}}$$

$$= \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

Form of Cost Function for Logistic Regression

Cost

$$\overset{\text{cost}}{J(\vec{w}, b)} = \frac{1}{m}\sum_{i=1}^{m} L\underbrace{\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)}_{\text{loss}}$$

$$= \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

convex
↳ can reach a global minimum

Form of Cost Function for Logistic Regression

Cost

cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m} L\Big(\underbrace{f_{\vec{w},b}\big(\vec{x}^{(i)}\big), y^{(i)}}_{\text{loss}}\Big)$$

$$= \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

convex
↳ can reach a global minimum

find w, b that minimize cost J

Table of Contents

Simplified Loss Function

Simplified loss function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$\boxed{L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)}$$

Simplified Loss Function

Simplified loss function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} \boxed{-\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)} & \text{if } y^{(i)} = 1 \\ -\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 0 \end{cases}$$

$$\boxed{L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -\underset{1}{y^{(i)}}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) \cancel{- \underbrace{\overset{(1-1)}{\left(1 - y^{(i)}\right)}}_{0}\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)}}$$

if $y^{(i)} = 1$:

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -1 \log\left(f(\vec{x})\right)$$

Simplified Loss Function

Simplified loss function

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = \begin{cases} -\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) & \text{if } y^{(i)} = 1 \\ \boxed{-\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)} & \text{if } y^{(i)} = 0 \end{cases}$$

$$\boxed{L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -\underset{0}{\cancel{y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)}} - \underset{(1-0)}{\left(1 - y^{(i)}\right)}\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)}$$

if $y^{(i)} = 1$:

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -1 \log\left(f(\vec{x})\right)$$

if $y^{(i)} = 0$:

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -(1-0)\log\left(1-f(\vec{x})\right)$$

Simplified Loss Function

Simplified cost function

loss

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$$

cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}\left[L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)\right]$$

Simplified Loss Function

Simplified cost function

loss

$$L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right) = -y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) - \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)$$

cost

$$J(\vec{w}, b) = \frac{1}{m}\sum_{i=1}^{m}\left[L\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right), y^{(i)}\right)\right]$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - f_{\vec{w},b}\left(\vec{x}^{(i)}\right)\right)\right]$$

Table of Contents

Table of Contents

Gradient Descent for Logistic Regression

Training logistic regression

Find $\vec{w}, b$

Given new $\vec{x}$, output $f_{\vec{w},b}(\vec{x}) = \frac{1}{1+e^{-(\vec{w}\cdot\vec{x}+b)}}$

Gradient Descent for Logistic Regression

Gradient descent

$$J(\vec{\mathrm{w}}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)})\right) + (1 - y^{(i)})\log\left(1 - f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}^{(i)})\right)\right]$$

repeat {

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{\mathrm{w}}, b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{\mathrm{w}}, b)$$

}

Table of Contents

Partial Derivatives of Loss Function

Consider the computations for model prediction and error (i.e., loss) calculation *for a single data point* as follows:

Partial Derivatives of Loss Function

Consider the computations for model prediction and error (i.e., loss) calculation *for a single data point* as follows:

1. $z = w^T x + b$

Partial Derivatives of Loss Function

Consider the computations for model prediction and error (i.e., loss) calculation *for a single data point* as follows:

1. $z = w^T x + b$
2. $\hat{y} = a = \sigma(z)$

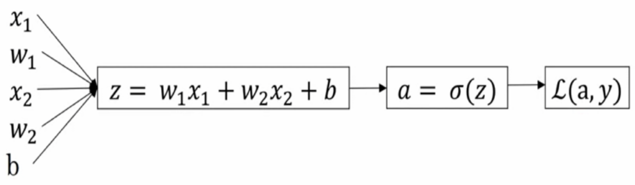
Partial Derivatives of Loss Function

Consider the computations for model prediction and error (i.e., loss) calculation *for a single data point* as follows:

1. $z = w^T x + b$
2. $\hat{y} = a = \sigma(z)$
3. $L(a, y) = -(y log(a) + (1 - y) log(1 - a))$

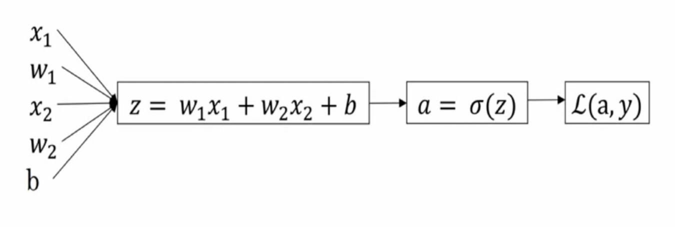
Partial Derivatives of Loss Function

We can model these computations for a single data point as a left-to-right stepwise progression through a computation graph as follows:



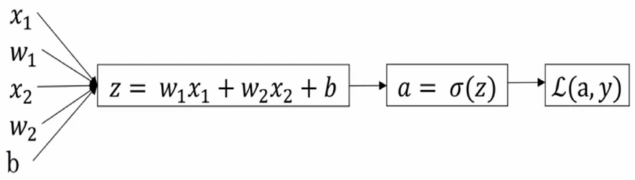
This computation process is called forward propagation.

Partial Derivatives of Loss Function



Now, we would like to find the following partial derivatives (in order to execute gradient descent):

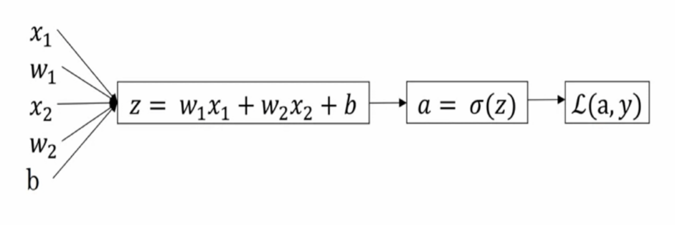
Partial Derivatives of Loss Function



Now, we would like to find the following partial derivatives (in order to execute gradient descent):

$$\frac{\partial L}{\partial w_1},$$

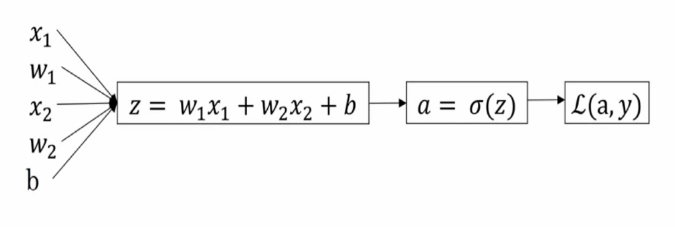
Partial Derivatives of Loss Function



Now, we would like to find the following partial derivatives (in order to execute gradient descent):

$$\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2},$$

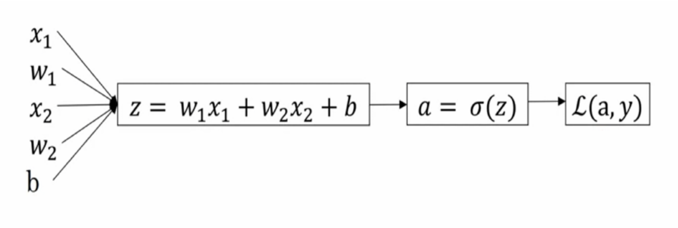
Partial Derivatives of Loss Function



Now, we would like to find the following partial derivatives (in order to execute gradient descent):

$$\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \text{ and } \frac{\partial L}{\partial b}$$

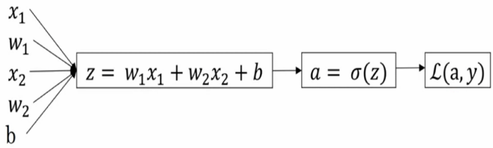
Partial Derivatives of Loss Function



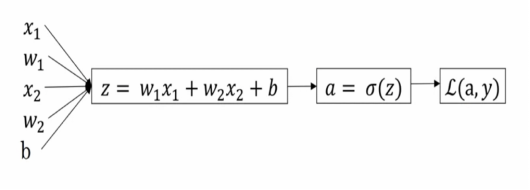
$$\text{Compute:} \qquad \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \text{ and } \frac{\partial L}{\partial b}$$

Approach: **Backpropagation**: Compute partial derivatives using right-to-left stepwise progression through this computation graph using the chain rule (from Calculus).

Partial Derivatives of Loss Function

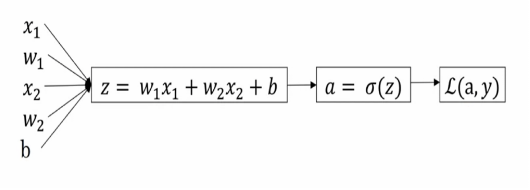


Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

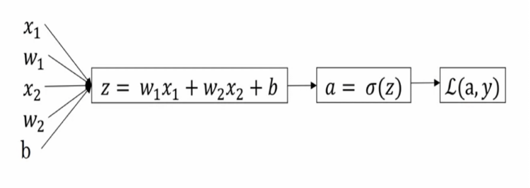
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$

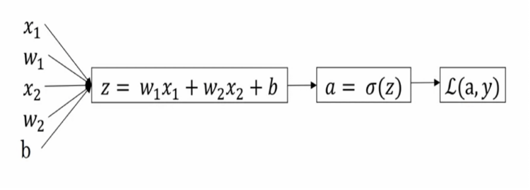
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$

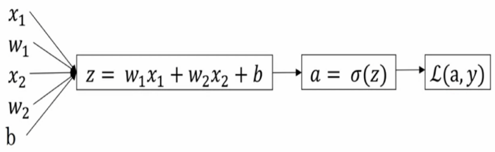
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$
3. $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{da}{dz}$ (Chain Rule)

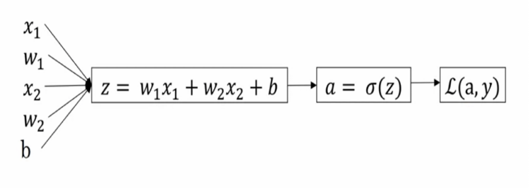
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$
3. $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{da}{dz}$ (Chain Rule)
4. $\frac{\partial z}{\partial w_1}, \frac{\partial z}{\partial w_2}, \frac{\partial z}{\partial b}$

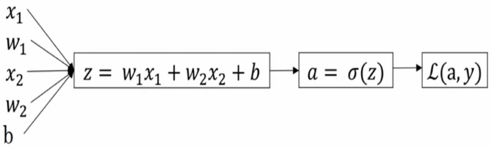
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$
3. $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{da}{dz}$ (Chain Rule)
4. $\frac{\partial z}{\partial w_1}, \frac{\partial z}{\partial w_2}, \frac{\partial z}{\partial b}$
5. $\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1}$ (Chain Rule)

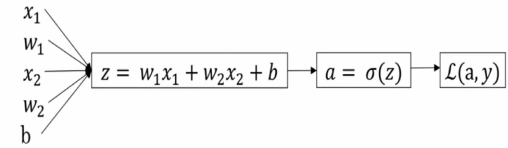
Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$
3. $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{da}{dz}$ (Chain Rule)
4. $\frac{\partial z}{\partial w_1}, \frac{\partial z}{\partial w_2}, \frac{\partial z}{\partial b}$
5. $\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1}$ (Chain Rule)
6. $\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_2}$ (Chain Rule)

Partial Derivatives of Loss Function



Backpropagation: Compute the following in order:

1. $\frac{\partial L}{\partial a}$
2. $\frac{\partial a}{\partial z}$
3. $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{da}{dz}$ (Chain Rule)
4. $\frac{\partial z}{\partial w_1}, \frac{\partial z}{\partial w_2}, \frac{\partial z}{\partial b}$
5. $\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1}$ (Chain Rule)
6. $\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_2}$ (Chain Rule)
7. $\frac{\partial L}{\partial b} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial b}$ (Chain Rule)

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

$$L(a, y) = -\left[y \log a + (1 - y) \log(1 - a)\right]$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

$$L(a, y) = -\left[y \log a + (1 - y) \log(1 - a)\right]$$
$$\therefore \frac{\partial L}{\partial a} = -\left[y \cdot \frac{1}{a} + (1 - y) \cdot \frac{-1}{(1 - a)}\right]$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

$$L(a, y) = -\left[y \log a + (1 - y) \log(1 - a)\right]$$

$$\therefore \frac{\partial L}{\partial a} = -\left[y \cdot \frac{1}{a} + (1 - y) \cdot \frac{-1}{(1 - a)}\right]$$

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

$$\begin{aligned}
L(a, y) &= -\left[y \log a + (1-y)\log(1-a)\right] \\
\therefore \frac{\partial L}{\partial a} &= -\left[y \cdot \frac{1}{a} + (1-y) \cdot \frac{-1}{(1-a)}\right] \\
&= -\left[\frac{y}{a} + \frac{(1-y)}{(1-a)}(-1)\right] \\
&= -\left[\frac{y}{a} - \frac{(1-y)}{(1-a)}\right]
\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

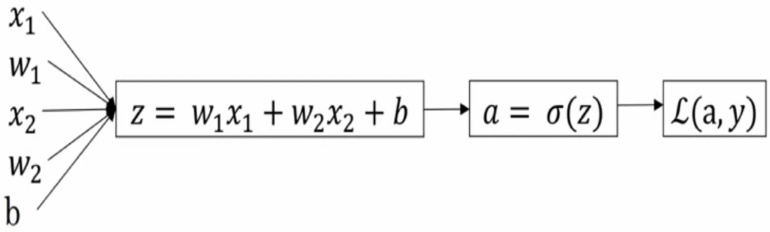
$$\begin{aligned}
L(a, y) &= -\left[y \log a + (1-y) \log(1-a)\right] \\
\therefore \frac{\partial L}{\partial a} &= -\left[y \cdot \frac{1}{a} + (1-y) \cdot \frac{-1}{(1-a)}\right] \\
&= -\left[\frac{y}{a} + \frac{(1-y)}{(1-a)}(-1)\right] \\
&= -\left[\frac{y}{a} - \frac{(1-y)}{(1-a)}\right] \\
&= -\frac{y}{a} + \frac{1-y}{1-a}
\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial a}$:

$$\begin{aligned}
L(a, y) &= -\left[y\log a + (1-y)\log(1-a)\right] \\
\therefore \frac{\partial L}{\partial a} &= -\left[y \cdot \frac{1}{a} + (1-y) \cdot \frac{-1}{(1-a)}\right] \\
&= -\left[\frac{y}{a} + \frac{(1-y)}{(1-a)}(-1)\right] \\
&= -\left[\frac{y}{a} - \frac{(1-y)}{(1-a)}\right] \\
&= -\frac{y}{a} + \frac{1-y}{1-a} \\
&= \frac{1-y}{1-a} - \frac{y}{a}
\end{aligned}$$

Partial Derivatives of Loss Function



Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule.

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$.

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$. Then, $du = 0$ and $dv = e^{-z}(-1) = -e^{-z}$.

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$. Then, $du = 0$ and $dv = e^{-z}(-1) = -e^{-z}$. So, we have:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$. Then, $du = 0$ and $dv = e^{-z}(-1) = -e^{-z}$. So, we have:

$$\frac{\partial a}{\partial z} = \frac{d\sigma(z)}{dz} = \frac{u'v - uv'}{v^2}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$. Then, $du = 0$ and $dv = e^{-z}(-1) = -e^{-z}$. So, we have:

$$\begin{aligned}\frac{\partial a}{\partial z} = \frac{d\sigma(z)}{dz} &= \frac{u'v - uv'}{v^2} \\ &= \frac{0 \cdot (1+e^{-z}) - (1)(-e^{-z})}{(1+e^{-z})^2}\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$:

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

Use the Quotient Rule. Let $u = 1$ and $v = 1 + e^{-z}$. Then, $du = 0$ and $dv = e^{-z}(-1) = -e^{-z}$. So, we have:

$$\begin{aligned}
\frac{\partial a}{\partial z} = \frac{d\sigma(z)}{dz} &= \frac{u'v - uv'}{v^2} \\
&= \frac{0 \cdot (1 + e^{-z}) - (1)(-e^{-z})}{(1 + e^{-z})^2} \\
&= \frac{e^{-z}}{(1 + e^{-z})^2}
\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

$$
\begin{aligned}
a = \sigma(z) &= \frac{1}{1+e^{-z}} \\
\therefore a(1+e^{-z}) &= 1 \\
a + ae^{-z} &= 1
\end{aligned}
$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

$$\begin{aligned} a = \sigma(z) &= \frac{1}{1+e^{-z}} \\ \therefore a(1+e^{-z}) &= 1 \\ a + ae^{-z} &= 1 \\ ae^{-z} &= 1 - a \end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)Note the following:

$$
\begin{aligned}
a = \sigma(z) &= \frac{1}{1+e^{-z}} \\
\therefore a(1+e^{-z}) &= 1 \\
a + ae^{-z} &= 1 \\
ae^{-z} &= 1 - a \\
e^{-z} &= \frac{1-a}{a}
\end{aligned}
$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) Also, note the following:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) Also, note the following:

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) Also, note the following:

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

$$\therefore a^2 = \left(\frac{1}{1+e^{-z}}\right)^2 = \frac{1}{(1+e^{-z})^2}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

$$e^{-z} = \frac{1-a}{a} \text{ and}$$
$$a^2 = \frac{1}{(1+e^{-z})^2}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.)   So we have:

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

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial a}{\partial z}$ (contd.) So we have:

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

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

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

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

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} = \left(\frac{1-y}{1-a} - \frac{y}{a}\right) \cdot a(1-a)$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$\begin{aligned}\frac{\partial L}{\partial z} &= \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} = \left(\frac{1-y}{1-a} - \frac{y}{a}\right) \cdot a(1-a) \\ &= \left[\frac{(1-y)a - y(1-a)}{a(1-a)}\right] \cdot a(1-a)\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$\begin{aligned}\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} &= \left(\frac{1-y}{1-a} - \frac{y}{a}\right) \cdot a(1-a) \\ &= \left[\frac{(1-y)a - y(1-a)}{a(1-a)}\right] \cdot a(1-a) \\ &= (1-y)a - y(1-a)\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$
\begin{aligned}
\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} &= \left(\frac{1-y}{1-a} - \frac{y}{a}\right) \cdot a(1-a) \\
&= \left[\frac{(1-y)a - y(1-a)}{a(1-a)}\right] \cdot a(1-a) \\
&= (1-y)a - y(1-a) \\
&= a - ay - y + ay
\end{aligned}
$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial z}$.

$$\begin{aligned}
\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z} &= \left(\frac{1-y}{1-a} - \frac{y}{a}\right) \cdot a(1-a) \\
&= \left[\frac{(1-y)a - y(1-a)}{a(1-a)}\right] \cdot a(1-a) \\
&= (1-y)a - y(1-a) \\
&= a - ay - y + ay \\
&= a - y
\end{aligned}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial w_1}$.

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial w_1}$.

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1}$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial w_1}$.

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1} \quad \text{where} \quad z = w_1 x_1 + w_2 x_2 + b$$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial w_1}$.

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1} \qquad \text{where} \qquad z = w_1 x_1 + w_2 x_2 + b$$

Now, $\dfrac{\partial z}{\partial w_1} = x_1,\ \dfrac{\partial z}{\partial w_2} = x_2$ and $\dfrac{\partial z}{\partial b} = 1$

Partial Derivatives of Loss Function

Derivation of $\frac{\partial L}{\partial w_1}$.

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial w_1} \quad \text{where} \quad z = w_1 x_1 + w_2 x_2 + b$$

$$\text{Now, } \frac{\partial z}{\partial w_1} = x_1, \ \frac{\partial z}{\partial w_2} = x_2 \text{ and } \frac{\partial z}{\partial b} = 1$$

$$\therefore \frac{\partial L}{\partial w_1} = (a - y)x_1, \ \frac{\partial L}{\partial w_2} = (a - y)x_2, \ \frac{\partial L}{\partial b} = a - y$$

Partial Derivatives of Loss Function

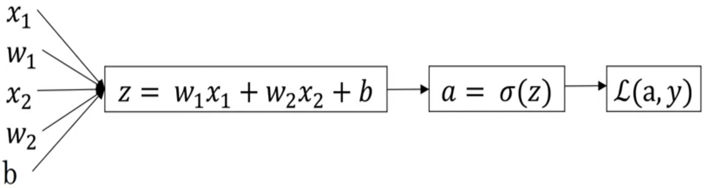


Table of Contents

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

But we really need

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

But we really need $\frac{\partial J}{\partial w_1}$,

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

But we really need $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$,

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

But we really need $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$

Partial Derivatives of Cost Function

Now, we have (for a single data sample):

- $\frac{\partial L}{\partial w_1} = (a - y)x_1$
- $\frac{\partial L}{\partial w_2} = (a - y)x_2$
- $\frac{\partial L}{\partial b} = a - y$

But we really need $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$ to run gradient descent on the entire training set.

Partial Derivatives of Cost Function

Compute $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$:

Partial Derivatives of Cost Function

Compute $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$:

$$J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

Partial Derivatives of Cost Function

Compute $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$:

$$J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

$$\therefore \frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial w_j} L(\hat{y}^{(i)}, y^{(i)}) \quad j = 1, \ldots, n$$

Partial Derivatives of Cost Function

Compute $\frac{\partial J}{\partial w_1}$, $\frac{\partial J}{\partial w_2}$, and $\frac{\partial J}{\partial b}$:

$$J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$$

$$\therefore \frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial w_j} L(\hat{y}^{(i)}, y^{(i)}) \quad j = 1, \ldots, n$$

$$\text{and} \quad \frac{\partial}{\partial b} J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial b} L(\hat{y}^{(i)}, y^{(i)})$$

Partial Derivatives of Cost Function

Since $\frac{\partial L}{\partial w_j} = (a - y)x_j$ for $j = 1, \ldots, n$ and $\frac{\partial L}{\partial b} = a - y$, we have, by substitution:

Partial Derivatives of Cost Function

Since $\frac{\partial L}{\partial w_j} = (a - y)x_j$ for $j = 1, \ldots, n$ and $\frac{\partial L}{\partial b} = a - y$, we have, by substitution:

$$\frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m}\sum_{i=1}^{m}(a^{(i)} - y^{(i)})x_j^{(i)} \quad j = 1, \ldots, n$$

Partial Derivatives of Cost Function

Since $\frac{\partial L}{\partial w_j} = (a - y)x_j$ for $j = 1, \ldots, n$ and $\frac{\partial L}{\partial b} = a - y$, we have, by substitution:

$$\frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m} \sum_{i=1}^{m} (a^{(i)} - y^{(i)}) x_j^{(i)} \quad j = 1, \ldots, n$$

$$\text{and} \quad \frac{\partial}{\partial b} J_{\vec{w},b}(\vec{x}) = \frac{1}{m} \sum_{i=1}^{m} (a^{(i)} - y^{(i)})$$

Partial Derivatives of Cost Function

Since $a^{(i)} = f_{\vec{w},b}(\vec{x}^{(i)})$ for $i = 1, \ldots, m$ we have, again, by substitution:

Partial Derivatives of Cost Function

Since $a^{(i)} = f_{\vec{w},b}(\vec{x}^{(i)})$ for $i = 1, \ldots, m$ we have, again, by substitution:

$$\frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

Partial Derivatives of Cost Function

Since $a^{(i)} = f_{\vec{w},b}(\vec{x}^{(i)})$ for $i = 1, \ldots, m$ we have, again, by substitution:

$$\frac{\partial}{\partial w_j} J_{\vec{w},b}(\vec{x}) = \frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

and $$\frac{\partial}{\partial b} J_{\vec{w},b}(\vec{x}) = \frac{1}{m} \sum_{i=1}^{m} (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})$$

Table of Contents

Gradient Descent Implementation

Gradient descent

$$J(\vec{w},b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right]$$

repeat {

$j=1...n$

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w},b) \qquad \frac{\partial}{\partial w_j} J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{w},b)$$

}

Gradient Descent Implementation

Gradient descent

$$J(\vec{\mathrm{w}}, b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right)\right)\right]$$

repeat {

$j=1\ldots n$

$$w_j = w_j - \alpha\frac{\partial}{\partial w_j}J(\vec{\mathrm{w}}, b) \qquad \frac{\partial}{\partial w_j}J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

$$b = b - \alpha\frac{\partial}{\partial b}J(\vec{\mathrm{w}}, b) \qquad \frac{\partial}{\partial b}J(\vec{\mathrm{w}}, b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)$$

}

Gradient Descent Implementation

Gradient descent

$$J(\vec{w},b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1-y^{(i)})\log\left(1-f_{\vec{w},b}(\vec{x}^{(i)})\right)\right]$$

repeat {

$j = 1 \ldots n$

$$w_j = w_j - \alpha \underbrace{\frac{\partial}{\partial w_j}J(\vec{w},b)} \qquad \frac{\partial}{\partial w_j}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)x_j^{(i)}$$

$$b = b - \alpha \underbrace{\frac{\partial}{\partial b}J(\vec{w},b)} \qquad \frac{\partial}{\partial b}J(\vec{w},b) = \frac{1}{m}\sum_{i=1}^{m}\left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$$

} simultaneous updates

Gradient Descent Implementation

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)} \right) \right]$$

} simultaneous updates

Linear regression $f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b$

Logistic regression $f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = \dfrac{1}{1 + e^{-(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b)}}$

Gradient Descent Implementation

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

Same concepts:

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$

Gradient Descent Implementation

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

Same concepts:

- Monitor gradient descent (learning curve)

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$

Gradient Descent Implementation

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

Same concepts:

- Monitor gradient descent (learning curve)
- Vectorized implementation

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \dfrac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$

Gradient Descent Implementation

repeat {

$$w_j = w_j - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) x_j^{(i)} \right]$$

$$b = b - \alpha \left[\frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right) \right]$$

} simultaneous updates

Same concepts:

- Monitor gradient descent (learning curve)
- Vectorized implementation
- Feature scaling

Linear regression $f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b$

Logistic regression $f_{\vec{w},b}(\vec{x}) = \frac{1}{1 + e^{-(\vec{w} \cdot \vec{x} + b)}}$