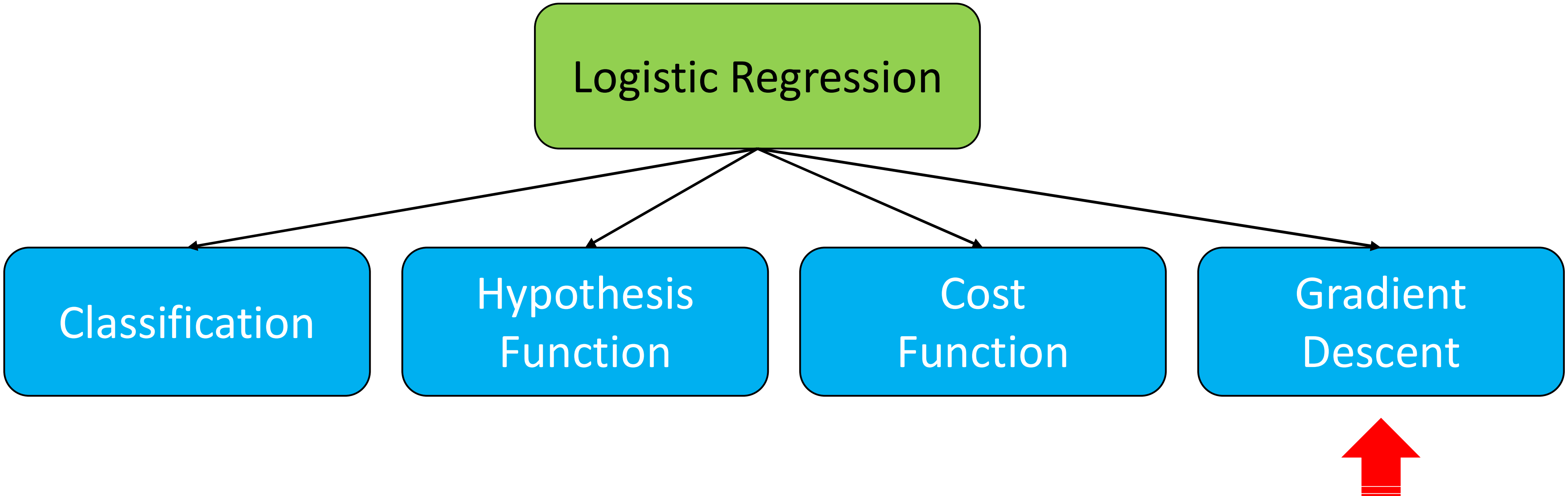
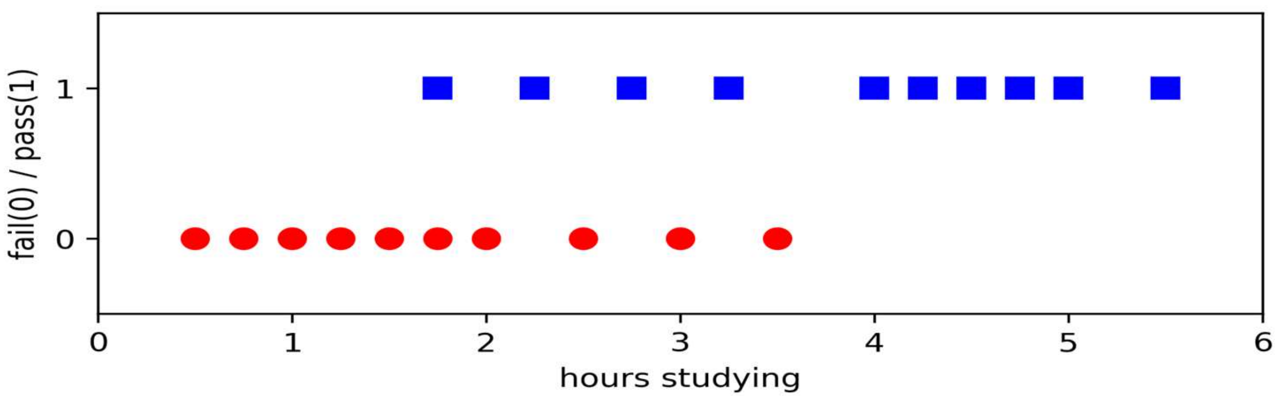
Logistic Regression

Outline


Logistic Regression
Classification
Hypothesis Function
Cost Function
Gradient Descent

Binary outcomes are common and important

- The patient survives the operation, or does not.
- The accused is convicted, or is not.
- The customer makes a purchase, or does not.
- The marriage lasts at least five years, or does not.
- The student graduates, or does not.



Categorical Response Variables

Whether or not a person smokes

Binary Response

$$Y = \begin{cases} \text{Non} - \text{smoker} \\ \text{Smoker} \end{cases}$$

Success of a medical treatment

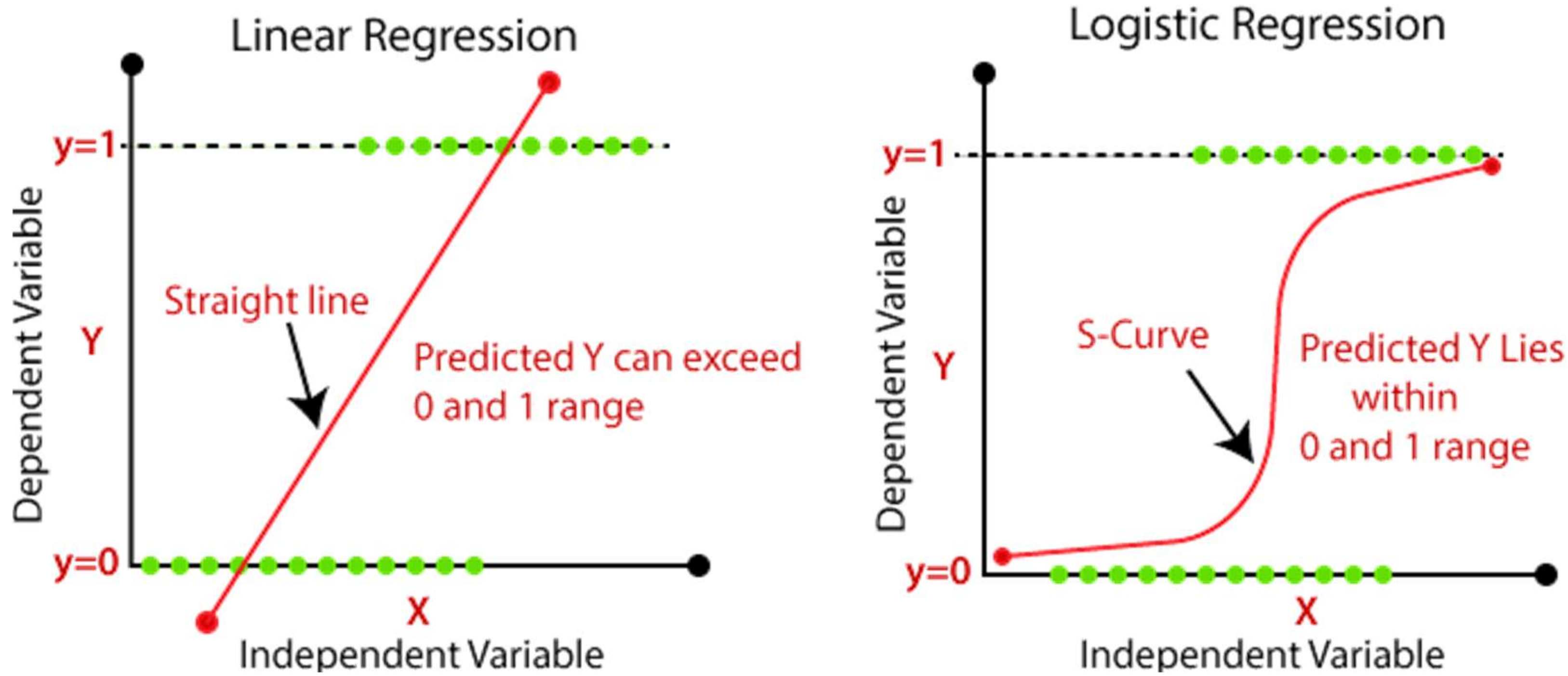
$$Y = \begin{cases} \text{Survives} \\ \text{Dies} \end{cases}$$

Opinion poll responses

Ordinal Response

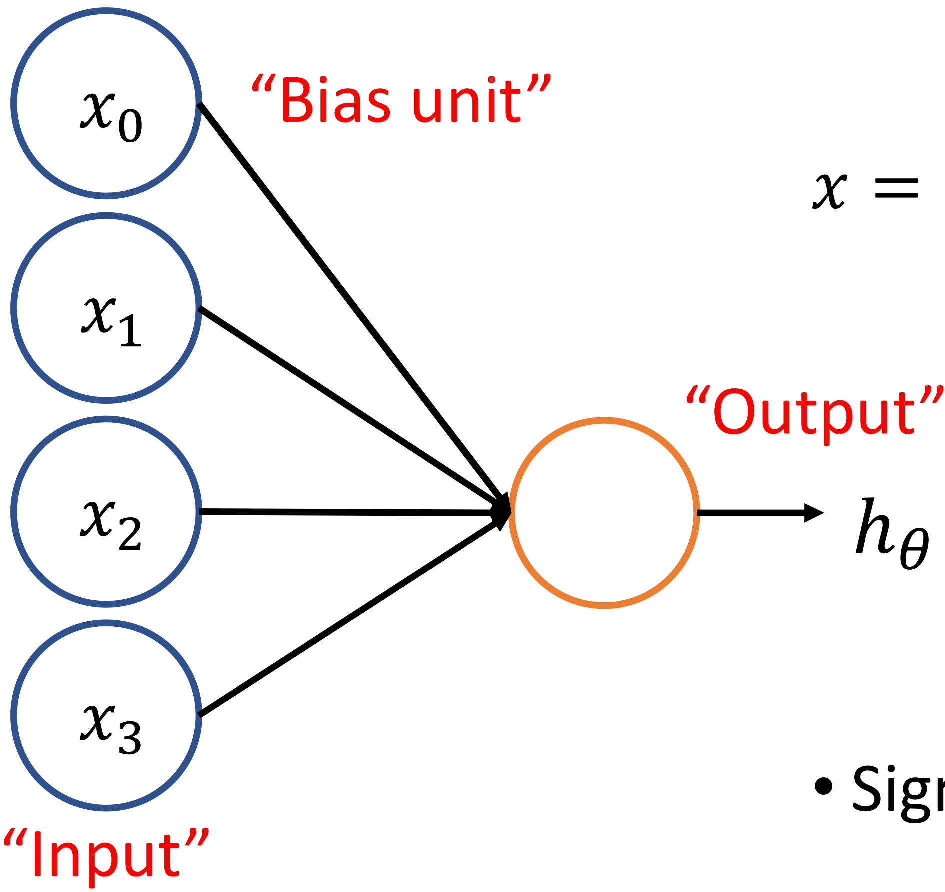
$$Y = \begin{cases} \text{Agree} \\ \text{Neutral} \\ \text{Disagree} \end{cases}$$

Difference between linear regression and logistic regression



https://www.kaggle.com/

Sigmoid Function



$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

“Weights”
“Parameters”

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^\top x}}$$

- Sigmoid (logistic) activation function

Slide credit: Andrew Ng

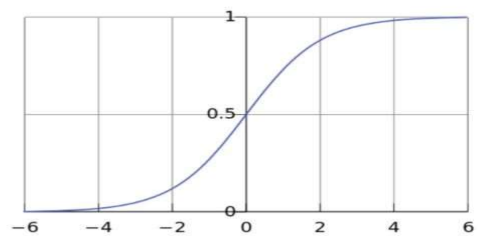
Learning a Logistic Regression Model

- How to learn a *logistic regression model* $\boldsymbol{h_\theta}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{x})$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$ and $\boldsymbol{x} = [\boldsymbol{x_0}, \dots, \boldsymbol{x_m}]$?
  - By minimizing the following cost function:

$$\text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x}), \text{y}) = -y\log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right) - (1-y)\log\left(1-\frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right)$$

  - That is:

$$\underset{\theta}{\text{minimize}}\frac{1}{n}\sum_{i=1}^{n}\text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x})^{(i)}, y^{(i)})$$



$$\equiv$$

$$\underset{\theta}{\text{minimize}}\boxed{\frac{1}{n}\sum_{i=1}^{n}-y^{(i)}\log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right) - (1-y)\log\left(1-\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right)}$$

Cost function
$\boldsymbol{J(\theta)}$

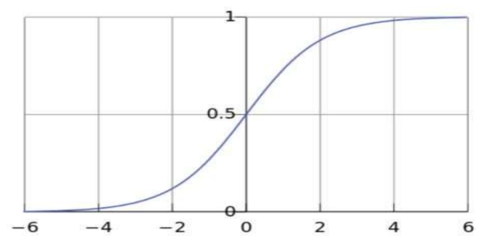
Learning a Logistic Regression Model

- How to learn a *logistic regression model* $\boldsymbol{h_\theta}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{\theta^T x})$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \ldots, \boldsymbol{\theta_m}]$ and $\boldsymbol{x} = [\boldsymbol{x_0}, \ldots, \boldsymbol{x_m}]$?
  - By minimizing the following cost function:

$$\text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x}), \text{y}) = -y \log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right) - (1-y)\log\left(1 - \frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right)$$

  - That is:

$$\underset{\theta}{\text{minimize}} \frac{1}{n}\sum_{i=1}^{n} \text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x})^{(i)}, y^{(i)})$$



$$\equiv$$

$$\underset{\theta}{\text{minimize}} \boxed{\frac{1}{n}\sum_{i=1}^{n} -y^{(i)} \log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right) - (1-y)\log\left(1 - \frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right)}$$

Cost function
$\boldsymbol{J(\theta)}$

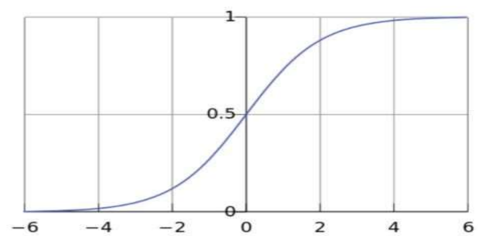
Learning a Logistic Regression Model

- How to learn a *logistic regression model* $\boldsymbol{h_\theta}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{x})$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$ and $\boldsymbol{x} = [\boldsymbol{x_0}, \dots, \boldsymbol{x_m}]$?
  - By minimizing the following cost function:

$$\text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x}), \text{y}) = -y \log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right) - (1-y)\log\left(1 - \frac{1}{1+e^{-\boldsymbol{\theta}^T x}}\right)$$

  - That is:

$$\underset{\theta}{\text{minimize}} \frac{1}{n}\sum_{i=1}^{n} \text{Cost}(\boldsymbol{h_\theta}(\boldsymbol{x})^{(i)}, y^{(i)})$$



$$\equiv$$

$$\underset{\theta}{\text{minimize}} \boxed{\frac{1}{n}\sum_{i=1}^{n} -y^{(i)} \log\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right) - (1-y)\log\left(1 - \frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}}\right)}$$

Cost function
$\boldsymbol{J(\theta)}$

Gradient Descent For Logistic Regression

- **Outline**:
  - Have cost function $\boldsymbol{J(\theta)}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$
  - Start off with some guesses for $\theta_0, \dots, \theta_m$
    - It does not really matter what values you start off with, but a common choice is to set them all initially to zero
  - Repeat until convergence{

Partial derivative

$$\theta_j = \theta_j - \alpha \frac{\partial \boldsymbol{J(\theta)}}{\partial \theta_j}$$

Note: **Update all θ_j simulatenously**

}

***Learing rate**, which controls how big a step we take when we update θ_j*

Gradient Descent For Logistic Regression

- **Outline**:
  - Have cost function $\boldsymbol{J(\theta)}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$
  - Start off with some guesses for $\theta_0, \dots, \theta_m$
    - It does not really matter what values you start off with, but a common choice is to set them all initially to zero
  - Repeat until convergence{

$$\theta_j = \theta_j - \alpha \boxed{\sum_{i=1}^{n} \left(\frac{1}{1 + e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)} \right) x_j^{(i)}}$$

The final formula after applying partial derivatives

}

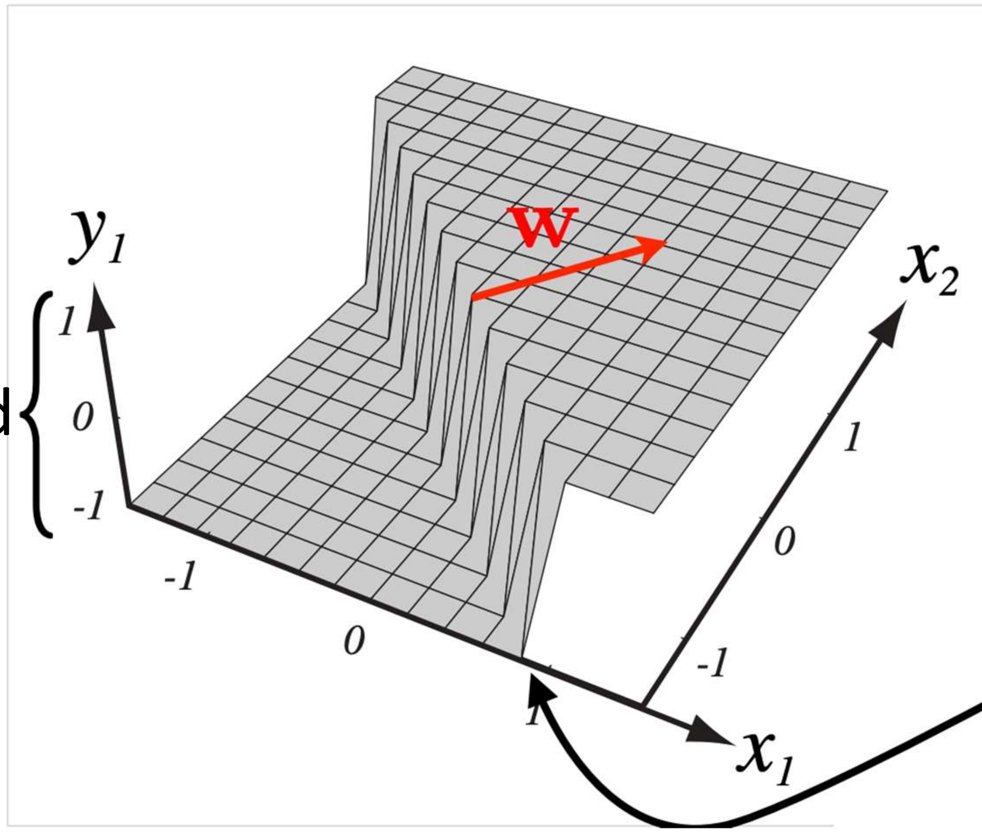
Inference After Learning

- After learning the parameters $\boldsymbol{\theta}$ = $[\theta_0, \dots, \theta_m]$, we can predict the output of any new unseen $\boldsymbol{x} = [\boldsymbol{x_0}, \dots, \boldsymbol{x_m}]$ as follows:

$$
\begin{cases}
\boldsymbol{if\ h_\theta(x)} = \dfrac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} < \mathbf{0.5\ predict\ 0} \\[2ex]
\boldsymbol{Else\ if\ h_\theta(x)} = \dfrac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} \geq \mathbf{0.5\ predict\ 1}
\end{cases}
$$

Visualization of weights, bias, activation function

range determined by g(.)

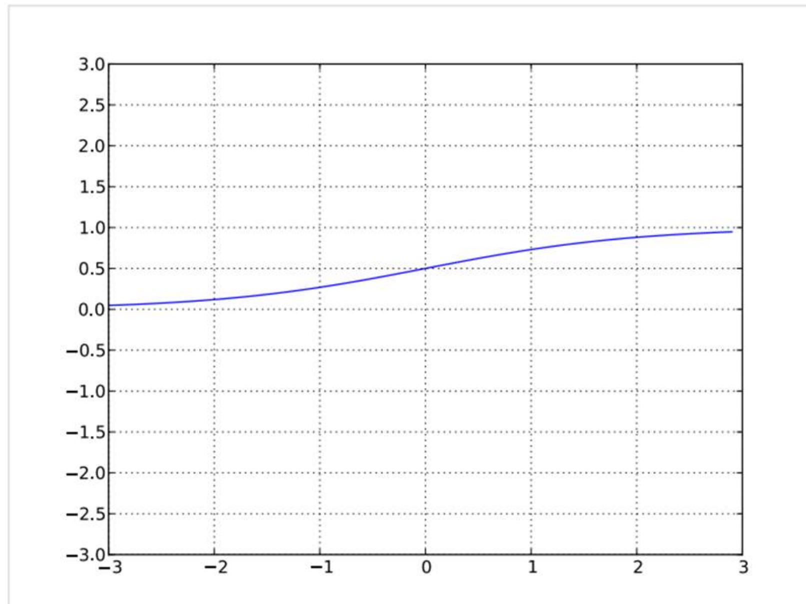


bias b only change the position of the hyperplane

Slide credit: Hugo Larochelle

Activation - sigmoid

- Squashes the neuron's pre-activation between 0 and 1
- Always positive
- Bounded
- Strictly increasing



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

Slide credit: Hugo Larochelle

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

	and	vaccine	the	of	nigeria	y
Email **a**	1	1	0	1	1	1
Email **b**	0	0	1	1	0	0
Email **c**	0	1	1	0	0	1
Email **d**	1	0	0	1	0	0
Email **e**	1	0	1	0	1	1
Email **f**	1	0	1	1	0	0

A Training Dataset

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

	and	vaccine	the	of	nigeria	y
Email **a**	1	1	0	1	1	1
Email **b**	0	0	1	1	0	0
Email **c**	0	1	1	0	0	1
Email **d**	1	0	0	1	0	0
Email **e**	1	0	1	0	1	1
Email **f**	1	0	1	1	0	0

1 entails that a word (i.e., “and”) is *present* in an email (i.e., “Email **a**”)

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming α = 0.5 and starting with θ = [0, 0, 0, 0, 0, 0]

	and	vaccine	the	of	nigeria	y
Email **a**	1	1	0	1	1	1
Email **b**	0	0	1	1	0	0
Email **c**	0	1	1	0	0	1
Email **d**	1	0	0	1	0	0
Email **e**	1	0	1	0	1	1
Email **f**	1	0	1	1	0	0

0 entails that a word (i.e., "and") is *abscent* in an email (i.e., "Email **b**")

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

We define 6 parameters (the first one, i.e., θ_0, is the intercept)

5 words (or ***features***) = $[\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{x_4}, \boldsymbol{x_5}]$

	$\boldsymbol{x_1}$ = and	$\boldsymbol{x_2}$ = vaccine	$\boldsymbol{x_3}$ = the	$\boldsymbol{x_4}$ = of	$\boldsymbol{x_5}$ = nigeria	y
Email **a**	1	1	0	1	1	1
Email **b**	0	0	1	1	0	0
Email **c**	0	1	1	0	0	1
Email **d**	1	0	0	1	0	0
Email **e**	1	0	1	0	1	1
Email **f**	1	0	1	1	0	0

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

The parameter vector:
$\boldsymbol{\theta} = [\boldsymbol{\theta_0}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \boldsymbol{\theta_4}, \boldsymbol{\theta_5}]$

x = $[\boldsymbol{x_0}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{x_4}, \boldsymbol{x_5}]$ *The feature vector*

	$x_0 = 1$	x_1 = and	x_2 = vaccine	x_3 = the	x_4 = of	x_5 = nigeria	y
Email **a**	1	1	1	0	1	1	1
Email **b**	1	0	0	1	1	0	0
Email **c**	1	0	1	1	0	0	1
Email **d**	1	1	0	0	1	0	0
Email **e**	1	1	0	1	0	1	1
Email **f**	1	1	0	1	1	0	0

To account for the intercept

Recap: Gradient Descent For Logistic Regression

- **Outline**:
  - Have cost function $\boldsymbol{J(\theta)}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$
  - Start off with some guesses for $\theta_0, \dots, \theta_m$
    - It does not really matter what values you start off with, but a common choice is to set them all initially to zero
  - Repeat until convergence{

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{n} \left(\frac{1}{1 + e^{-\boxed{\boldsymbol{\theta}^T x^{(i)}}}} - y^{(i)} \right) {x_j}^{(i)}$$

First, let us calculate this factor for every example in our training dataset

}

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0

Recap: Gradient Descent For Logistic Regression

- **Outline**:
  - Have cost function $\boldsymbol{J(\theta)}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \dots, \boldsymbol{\theta_m}]$
  - Start off with some guesses for $\theta_0, \dots, \theta_m$
    - It does not really matter what values you start off with, but a common choice is to set them all initially to zero
  - Repeat until convergence{

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{n} \boxed{\left(\frac{1}{1 + e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)} \right) x_j^{(i)}}$$

}

Second, let us calculate this equation for every example in our training dataset and for every $\boldsymbol{\theta_j}$, where *j* is between *0* and *m*

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$	$\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y}\right)\boxed{\boldsymbol{x}_0}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\theta^T x}} - \boldsymbol{y})\boldsymbol{x_0}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	$(\frac{1}{1+e^{-\mathbf{0}}} - \mathbf{1}) \times \mathbf{1}$ = -0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	$(\frac{1}{1+1} - \mathbf{0}) \times \mathbf{1}$ = 0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	$(\frac{1}{1+1} - \mathbf{1}) \times \mathbf{1}$ = -0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	$(\frac{1}{1+1} - \mathbf{0}) \times \mathbf{1}$ = 0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	$(\frac{1}{1+1} - \mathbf{1}) \times \mathbf{1}$ = -0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	$(\frac{1}{1+1} - \mathbf{0}) \times \mathbf{1}$ = 0.5

Recap: Gradient Descent For Logistic Regression

- **Outline**:
  - Have cost function $\boldsymbol{J(\theta)}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \ldots, \boldsymbol{\theta_m}]$
  - Start off with some guesses for $\theta_0, \ldots, \theta_m$
    - It does not really matter what values you start off with, but a common choice is to set them all initially to zero
  - Repeat until convergence{

$$\boxed{\theta_j = \theta_j - \alpha \sum_{i=1}^{n} \left(\frac{1}{1 + e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)} \right) x_j{}^{(i)}} \dashrightarrow \text{Third, let us compute every } \boldsymbol{\theta_j}$$

}

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x_0}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	-0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right){x_0}^{(i)} = \mathbf{0}$$

Then,

$$\theta_0 = \theta_0 - \alpha \times \mathbf{0}$$

New $\boldsymbol{\theta_0}$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T \boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x_0}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	-0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right){x_0}^{(i)} = \mathbf{0}$$

Then,

$$\theta_0 = \theta_0 - \alpha \times \mathbf{0}$$

Old $\boldsymbol{\theta_0}$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x}_0$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	-0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right) {x_0}^{(i)} = \mathbf{0}$$

Then,

$$\theta_0 = \theta_0 - \alpha \times \mathbf{0}$$

$$= 0 - 0.5 \times \mathbf{0} = \mathbf{0}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5]$$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boxed{\boldsymbol{x}_1}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x_1}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	0
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right) {x_1}^{(i)} = \mathbf{0}$$

Then,

$$\theta_1 = \theta_1 - \alpha \times \mathbf{0}$$

$$= 0 - 0.5 \times \mathbf{0} = \mathbf{0}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \boldsymbol{\theta_4}, \boldsymbol{\theta_5}]$$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T \boldsymbol{x}}} - \boldsymbol{y}\right)\boxed{\boldsymbol{x}_2}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x}_2$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	-0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	0
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right) {x_2}^{(i)} = -\mathbf{1}$$

Then,

$$\theta_2 = \theta_2 - \alpha \times (-\mathbf{1})$$

$$= 0 - 0.5 \times (-\mathbf{1}) = \mathbf{0.5}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \boldsymbol{\theta}_3, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5]$$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T \boldsymbol{x}}} - \boldsymbol{y})\boxed{\boldsymbol{x}_3}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T \boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x_3}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	0
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	-0.5
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right) {x_3}^{(i)} = \mathbf{0}$$

Then,

$$\theta_3 = \theta_3 - \alpha \times \mathbf{0}$$

$$= 0 - 0.5 \times 0 = \mathbf{0}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \boldsymbol{\theta_4}, \boldsymbol{\theta_5}]$$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$(\frac{1}{1+e^{-\theta^T x}} - \boldsymbol{y})\boxed{x_4}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$	$(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}} - \boldsymbol{y})\boldsymbol{x}_4$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0.5
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	0
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0.5
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	0
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0.5

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right){x_4}^{(i)} = \mathbf{1}$$

$\boldsymbol{Then,}$

$$\theta_4 = \theta_4 - \alpha \times \mathbf{1}$$

$$= 0 - 0.5 \times 1 = \mathbf{-0.5}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \boldsymbol{\theta}_5]$$

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming $\boldsymbol{\alpha}$ = 0.5 and starting with $\boldsymbol{\theta}$ = [0, 0, 0, 0, 0, 0]

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T\boldsymbol{x}$	$\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T\boldsymbol{x}}}-\boldsymbol{y}\right)\boxed{\boldsymbol{x}_5}$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	

A Concrete Example: The Training Phase

- Let us apply logistic regression on the spam email recognition problem, assuming α = 0.5 and starting with θ = [0, 0, 0, 0, 0, 0]

x	y	$\theta^T x$	$(\frac{1}{1+e^{-\theta^T x}} - y)x_5$
[1,1,1,0,1,1]	**1**	[0,0,0,0,0,0]×[1,1,1,0,1,1]=0	-0.5
[1,0,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,0,0,1,1,0]=0	0
[1,0,1,1,0,0]	**1**	[0,0,0,0,0,0]×[1,0,1,1,0,0]=0	0
[1,1,0,0,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,0,1,0]=0	0
[1,1,0,1,0,1]	**1**	[0,0,0,0,0,0]×[1,1,0,1,0,1]=0	-0.5
[1,1,0,1,1,0]	**0**	[0,0,0,0,0,0]×[1,1,0,1,1,0]=0	0

$$\sum_{i=1}^{n}\left(\frac{1}{1+e^{-\boldsymbol{\theta}^T x^{(i)}}} - y^{(i)}\right) {x_5}^{(i)} = \mathbf{-1}$$

Then,

$$\theta_5 = \theta_5 - \alpha \times (-\mathbf{1})$$

$$= 0 - 0.5 \times (-1) = \mathbf{0.5}$$

New Paramter Vector:

$$\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$$

A Concrete Example: Testing

- Let us now *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$
  - **Note**: Testing is typically done over a portion of the dataset that is not used during training, but rather kept only for testing the accuracy of the algorithm's predictions thus far
  - In this example, we will test over all the examples that we *used* during training, *just* for illustrative purposes

A Concrete Example: Testing

- Let us *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, -\mathbf{0.5}, \mathbf{0.5}]$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$
[1,1,1,0,1,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,1,0,1,1]=0.5
[1,0,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,0,0,1,1,0]=-0.5
[1,0,1,1,0,0]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,0,1,1,0,0]=0.5
[1,1,0,0,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,0,1,0]=-0.5
[1,1,0,1,0,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,0,1]=0.5
[1,1,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,1,0]=-0.5

A Concrete Example: Testing

- Let us *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, -\mathbf{0.5}, \mathbf{0.5}]$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta}^T \boldsymbol{x}$	$\boldsymbol{h_\theta}(\boldsymbol{x}) = (\frac{1}{1+e^{-\theta^T x}})$
[1,1,1,0,1,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,1,0,1,1]=0.5	0.622459331
[1,0,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,0,0,1,1,0]=-0.5	0.377540669
[1,0,1,1,0,0]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,0,1,1,0,0]=0.5	0.622459331
[1,1,0,0,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,0,1,0]=-0.5	0.377540669
[1,1,0,1,0,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,0,1]=0.5	0.622459331
[1,1,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,1,0]=-0.5	0.377540669

A Concrete Example: Testing

- Let us *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

$(\boldsymbol{if}\ \boldsymbol{h_\theta(x)} \geq \mathbf{0.5}, \boldsymbol{y'} = \mathbf{1}; \boldsymbol{else}\ \boldsymbol{y'} = \mathbf{0})$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$\boldsymbol{h_\theta(x)} = (\frac{1}{1+e^{-\theta^T x}})$	**Predicted Class (or y')**
[1,1,1,0,1,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,1,0,1,1]=0.5	0.622459331	
[1,0,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,0,0,1,1,0]=-0.5	0.377540669	
[1,0,1,1,0,0]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,0,1,1,0,0]=0.5	0.622459331	
[1,1,0,0,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,0,1,0]=-0.5	0.377540669	
[1,1,0,1,0,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,0,1]=0.5	0.622459331	
[1,1,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,1,0]=-0.5	0.377540669	

A Concrete Example: Testing

- Let us *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

$(\boldsymbol{if\ h_\theta(x) \geq 0.5, y' = 1; else\ y' = 0})$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$\boldsymbol{h_\theta(x)} = (\frac{1}{1+e^{-\theta^T x}})$	**Predicted Class (or y')**
[1,1,1,0,1,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,1,0,1,1]=0.5	0.622459331	**1**
[1,0,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,0,0,1,1,0]=-0.5	0.377540669	**0**
[1,0,1,1,0,0]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,0,1,1,0,0]=0.5	0.622459331	**1**
[1,1,0,0,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,0,1,0]=-0.5	0.377540669	**0**
[1,1,0,1,0,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,0,1]=0.5	0.622459331	**1**
[1,1,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,1,0]=-0.5	0.377540669	**0**

A Concrete Example: Testing

- Let us *test* logistic regression on the spam email recognition problem, using the just learnt $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

$(\boldsymbol{if\ h_{\theta}(x) \geq 0.5, y' = 1; else\ y' = 0})$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\theta^T x}$	$\boldsymbol{h_{\theta}(x)} = (\frac{1}{1+e^{-\theta^T x}})$	Predicted Class (or y')
[1,1,1,0,1,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,1,0,1,1]=0.5	0.622459331	1
[1,0,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,0,0,1,1,0]=-0.5	0.377540669	0
[1,0,1,1,0,0]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,0,1,1,0,0]=0.5	0.622459331	1
[1,1,0,0,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,0,1,0]=-0.5	0.377540669	0
[1,1,0,1,0,1]	**1**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,0,1]=0.5	0.622459331	1
[1,1,0,1,1,0]	**0**	[0,0,0.5,0,-0.5,0.5]×[1,1,0,1,1,0]=-0.5	0.377540669	0

NO Mispredictions!

A Concrete Example: Inference

- Let us infer whether a given new email, say, ***k*** = [1, 0, 1, 0, 0, 1] is a spam or not, using logistic regression with the just learnt parameter vector $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

	$x_0 = 1$	$x_1 =$ and	$x_2 =$ vaccine	$x_3 =$ the	$x_4 =$ of	$x_5 =$ nigeria	y
Email **a**	1	1	1	0	1	1	1
Email **b**	1	0	0	1	1	0	0
Email **c**	1	0	1	1	0	0	1
Email **d**	1	1	0	0	1	0	0
Email **e**	1	1	0	1	0	1	1
Email **f**	1	1	0	1	1	0	0

Our Training Dataset

A Concrete Example: Inference

- Let us infer whether a given new email, say, $\boldsymbol{k}$ = [1, 0, 1, 0, 0, 1] is a spam or not, using logistic regression with the just learnt parameter vector $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

	$x_0 = 1$	x_1 = and	x_2 = vaccine	x_3 = the	x_4 = of	x_5 = nigeria	y
Email **a**	1	1	1	0	1	1	1
Email **b**	1	0	0	1	1	0	0
Email **c**	1	0	1	1	0	0	1
Email **d**	1	1	0	0	1	0	0
Email **e**	1	1	0	1	0	1	1
Email **f**	1	1	0	1	1	0	0
Email **k**	1	0	1	0	0	1	**?**

A Concrete Example: Inference

- Let us infer whether a given new email, say, ***k*** = [1, 0, 1, 0, 0, 1] is a spam or not, using logistic regression with the just learnt parameter vector $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

$$\boldsymbol{h_\theta}(\boldsymbol{x}) = \frac{1}{1 + e^{-\boxed{\boldsymbol{\theta}^T \boldsymbol{x}}}} \qquad \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0.5} \\ \mathbf{0} \\ \mathbf{-0.5} \\ \mathbf{0.5} \end{bmatrix} [\mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{1}] = (\mathbf{0.5} \times \mathbf{1}) + (\mathbf{0.5} \times \mathbf{1}) = \mathbf{1}$$

$$= \frac{1}{1 + e^{-\mathbf{1}}}$$

$$= 0.731$$

$$\geq 0.5$$ ➔ Class 1 (i.e., Spam)

A Concrete Example: Inference

- Let us infer whether a given new email, say, ***k*** = [1, 0, 1, 0, 0, 1] is a spam or not, using logistic regression with the just learnt parameter vector $\boldsymbol{\theta} = [\mathbf{0}, \mathbf{0}, \mathbf{0.5}, \mathbf{0}, \mathbf{-0.5}, \mathbf{0.5}]$

	$x_0 = 1$	x_1 = and	x_2 = vaccine	x_3 = the	x_4 = of	x_5 = nigeria	y
Email **a**	1	1	1	0	1	1	1
Email **b**	1	0	0	1	1	0	0
Email **c**	1	0	1	1	0	0	1
Email **d**	1	1	0	0	1	0	0
Email **e**	1	1	0	1	0	1	1
Email **f**	1	1	0	1	1	0	0
Email **k**	1	0	1	0	0	1	**1**

Somehow interesting since it considered “vaccine” and “nigeria” indicative of spam!

Logistic Regression

- Sources:
  - https://www.kaggle.com/
  - http://research.cs.tamu.edu
  - http://web.iitd.ac.in
  - https://www3.nd.edu/