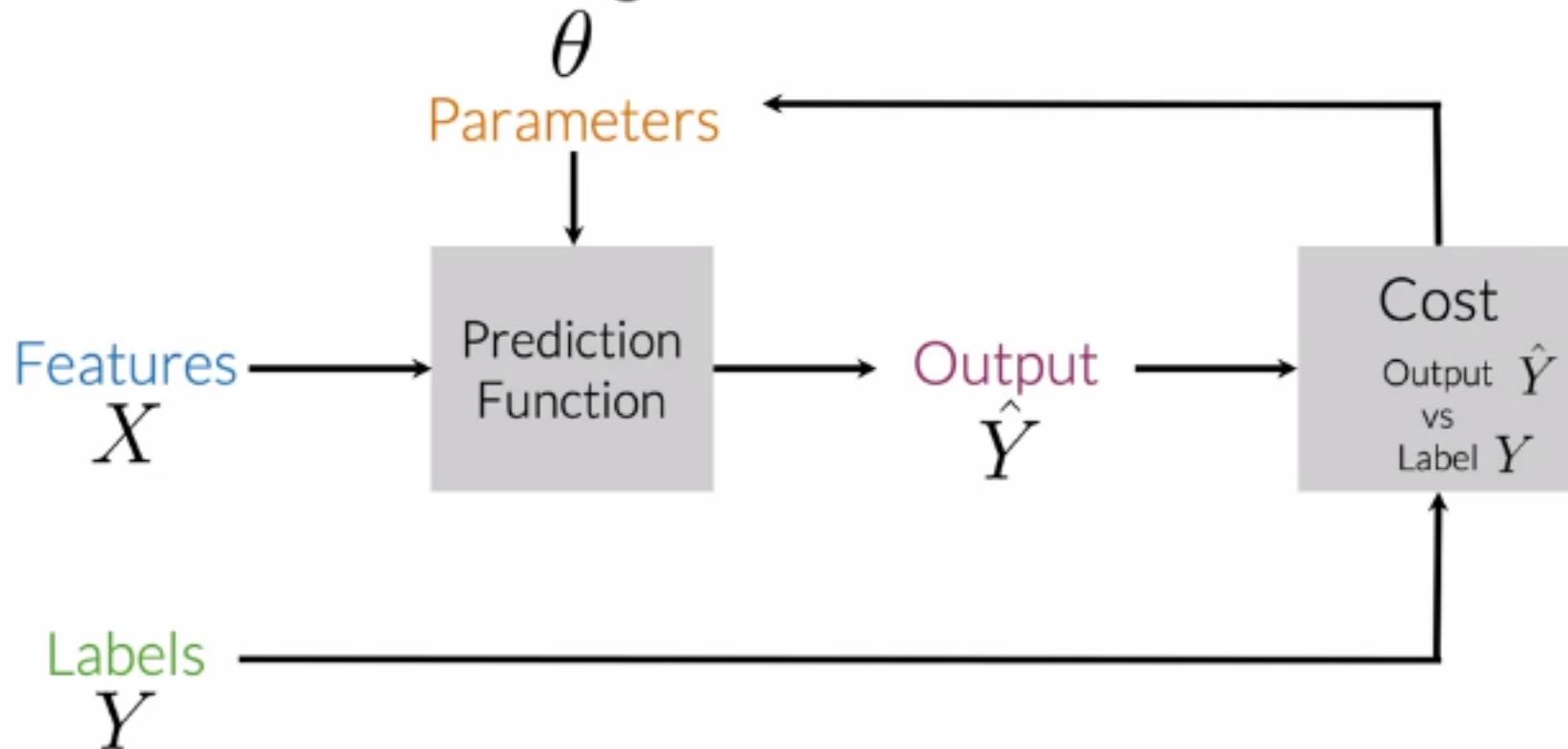
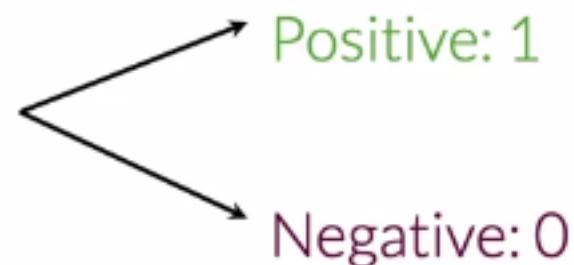


Supervised ML (training)

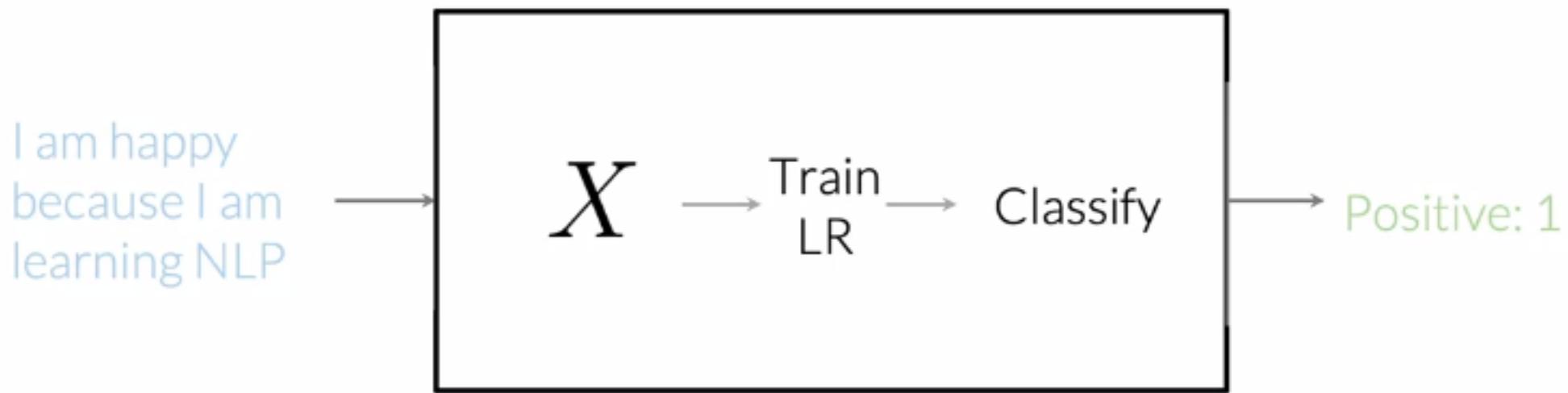


Sentiment analysis

Tweet: I am happy because I am learning NLP



Sentiment analysis



Vocabulary

Tweets:

[tweet_1, tweet_2, ..., tweet_m]



I am happy because I am learning NLP

...

I hated the movie

$V =$

[I, am, happy, because, learning, NLP, ... hated, the, movie]

Feature extraction

I am happy because I am learning NLP

[I, am, happy, because, learning, NLP, ... hated, the, movie]

↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
[1, 1, 1, 1, 1, 1, ... 0, 0, 0]

A lot of zeros! That's a sparse representation.

Problems with sparse representations

I am happy because I am learning NLP



Problems with sparse representations

I am happy because I am learning NLP

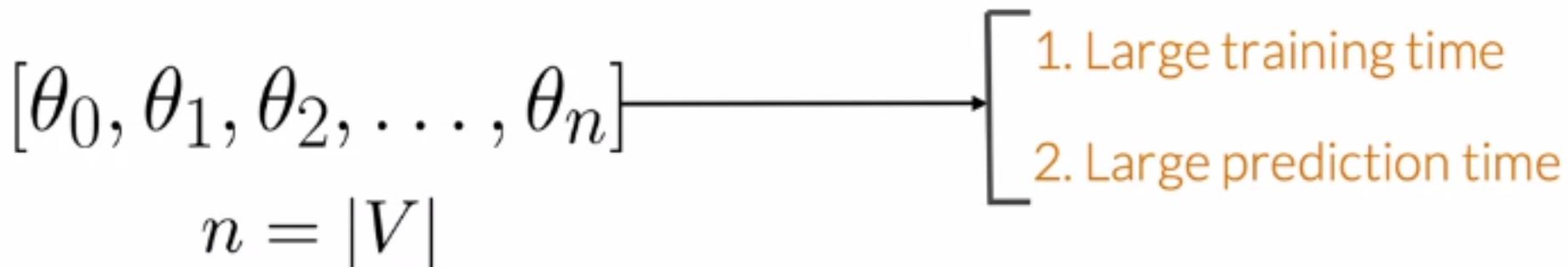


$$[\theta_0, \theta_1, \theta_2, \dots, \theta_n]$$

$$n = |V|$$

Problems with sparse representations

I am happy because I am learning NLP



Positive and negative counts

Corpus

I am happy because I am learning NLP

I am happy

I am sad, I am not learning NLP

I am sad

Positive and negative counts

Corpus

I am happy because I am learning NLP

I am happy

I am sad, I am not learning NLP

I am sad

Vocabulary

I

am

happy

because

learning

NLP

sad

not

Positive and negative counts

Positive tweets

I am happy because I am learning NLP

I am happy

Negative tweets

I am sad, I am not learning NLP

I am sad

Positive and negative counts

Positive tweets

I am happy because I am learning NLP

I am happy

Vocabulary

I

am

happy

because

learning

NLP

sad

not

Positive and negative counts

Positive tweets

I am happy because I am learning NLP

I am happy

| Vocabulary | PosFreq (1) |
|------------|-------------|
| I | 3 |
| am | 3 |
| happy | 2 |
| because | 1 |
| learning | 1 |
| NLP | 1 |
| sad | 0 |
| not | 0 |

Positive and negative counts

| Vocabulary | NegFreq (0) |
|------------|-------------|
| I | 3 |
| am | 3 |
| happy | 0 |
| because | 0 |
| learning | 1 |
| NLP | 1 |
| sad | 2 |
| not | 1 |

Negative tweets

I am sad, I am not learning NLP

I am sad

Word frequency in classes

| Vocabulary | PosFreq (1) | NegFreq (0) |
|------------|-------------|-------------|
| I | 3 | 3 |
| am | 3 | 3 |
| happy | 2 | 0 |
| because | 1 | 0 |
| learning | 1 | 1 |
| NLP | 1 | 1 |
| sad | 0 | 1 |
| not | 0 | 1 |

Word frequency in classes

| Vocabulary | PosFreq (1) | NegFreq (0) |
|------------|-------------|-------------|
| I | 3 | 3 |
| am | 3 | 3 |
| happy | 2 | 0 |
| because | 1 | 0 |
| learning | 1 | 1 |
| NLP | 1 | 1 |
| sad | 0 | 1 |
| not | 0 | 1 |

freqs: dictionary mapping from (word, class) to frequency

Word frequency in classes

| Vocabulary | PosFreq (1) | NegFreq (0) |
|------------|-------------|-------------|
| I | 3 | 3 |
| am | 3 | 3 |
| happy | 2 | 0 |
| because | 1 | 0 |
| learning | 1 | 1 |
| NLP | 1 | 1 |
| sad | 0 | 2 |
| not | 0 | 1 |

freqs: dictionary mapping from (word, class) to frequency

Feature extraction

freqs: dictionary mapping from (word, class) to frequency

Feature extraction

freqs: dictionary mapping from (word, class) to frequency

$$X_m = [\quad]$$


Features of
tweet m

Feature extraction

freqs: dictionary mapping from (word, class) to frequency

$$X_m = [1, \quad \quad \quad]$$

↓ ↓
Features of Bias
tweet m

Feature extraction

freqs: dictionary mapping from (word, class) to frequency

$$X_m = [1, \sum_w freqs(w, 1),]$$

↓ ↓ ↓
Features of Bias Sum Pos.
tweet m Frequencies

Feature extraction

freqs: dictionary mapping from (word, class) to frequency

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

↓ ↓ ↓ ↓

Features of Bias Sum Pos. Sum Neg.
tweet m Frequencies Frequencies

Feature extraction

I am sad, I am not learning NLP

Feature extraction

| Vocabulary | PosFreq (1) |
|------------|-------------|
| I | 3 |
| am | 3 |
| happy | 2 |
| because | 1 |
| learning | 1 |
| NLP | 1 |
| sad | 0 |
| not | 0 |

I am sad, I am not learning NLP

Feature extraction

| Vocabulary | PosFreq (1) |
|------------|-------------|
| I | 3 |
| am | 3 |
| happy | 2 |
| because | 1 |
| learning | 1 |
| NLP | 1 |
| sad | 0 |
| not | 0 |

I am sad, I am not learning NLP

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

Feature extraction

| Vocabulary | PosFreq (1) |
|------------|-------------|
| I | <u>3</u> |
| am | <u>3</u> |
| happy | 2 |
| because | 1 |
| learning | <u>1</u> |
| NLP | <u>1</u> |
| sad | <u>0</u> |
| not | <u>0</u> |

I am sad, I am not learning NLP

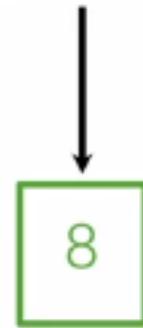
$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

Feature extraction

| Vocabulary | PosFreq (1) |
|------------|-------------|
| I | <u>3</u> |
| am | <u>3</u> |
| happy | 2 |
| because | 1 |
| learning | <u>1</u> |
| NLP | <u>1</u> |
| sad | <u>0</u> |
| not | <u>0</u> |

I am sad, I am not learning NLP

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$



Feature extraction

| Vocabulary | NegFreq (0) |
|------------|-------------|
| I | <u>3</u> |
| am | <u>3</u> |
| happy | 0 |
| because | 0 |
| learning | <u>1</u> |
| NLP | <u>1</u> |
| sad | <u>2</u> |
| not | <u>1</u> |

I am sad, I am not learning NLP

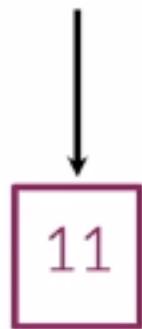
$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

Feature extraction

| Vocabulary | NegFreq (0) |
|------------|-------------|
| I | <u>3</u> |
| am | <u>3</u> |
| happy | 0 |
| because | 0 |
| learning | <u>1</u> |
| NLP | <u>1</u> |
| sad | <u>2</u> |
| not | <u>1</u> |

I am sad, I am not learning NLP

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$



Feature extraction

I am sad, I am not learning NLP

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

Feature extraction

I am sad, I am not learning NLP

$$X_m = [1, \sum_w freqs(w, 1), \sum_w freqs(w, 0)]$$

↓

$$X_m = [1, 8, 11]$$

Preprocessing: stop words and punctuation

@YMourri and @AndrewYNg are
tuning a GREAT AI model at
<https://deeplearning.ai!!!>

Preprocessing: stop words and punctuation

@YMourri and @AndrewYNg are
tuning a GREAT AI model at
[https://deeplearning.ai!!!](https://deeplearning.ai)

| Stop words | Punctuation |
|------------|-------------|
| and | , |
| is | . |
| are | : |
| at | ! |
| has | " |
| for | ' |
| a | |

Preprocessing: stop words and punctuation

@YMourri ~~and~~ @AndrewYNg ~~are~~
tuning ~~a~~ GREAT AI model ~~at~~
<https://deeplearning.ai/>!!!

Stop words
and
is
are
at
has
for
a

Punctuation
,

.

:

|

*

"

'

Preprocessing: stop words and punctuation

@YMourri ~~and~~ @AndrewYNg ~~are~~
tuning ~~a~~ GREAT AI model ~~at~~
[https://deeplearning.ai!!!](https://deeplearning.ai)

@YMourri @AndrewYNg tuning
GREAT AI model
[https://deeplearning.ai!!!](https://deeplearning.ai)

Stop words
and
is
are
at
has
for
a

Punctuation
,
.u
:
|
"
"

Preprocessing: stop words and punctuation

@YMourri @AndrewYNg tuning
GREAT AI model
<https://deeplearning.ai>!!!

| <u>Stop words</u> | <u>Punctuation</u> |
|-------------------|--------------------|
| and | , |
| is | . |
| a | : |
| at | ! |
| has | " |
| for | ' |
| of | |

Preprocessing: stop words and punctuation

@YMourri @AndrewYNg tuning
GREAT AI model
<https://deeplearning.ai>!!

| <u>Stop words</u> | <u>Punctuation</u> |
|-------------------|--------------------|
| and | , |
| is | . |
| a | : |
| at | ! |
| has | <u>a</u> |
| for | , |
| of | |

Preprocessing: stop words and punctuation

@YMurri @AndrewYNg tuning
GREAT AI model
<https://deeplearning.ai>!!!

@YMurri @AndrewYNg tuning
GREAT AI model
<https://deeplearning.ai>

| <u>Stop words</u> | <u>Punctuation</u> |
|-------------------|--------------------|
| and | , |
| is | . |
| a | : |
| at | ! |
| has | <u>a</u> |
| for | , |
| of | |

Preprocessing: Handles and URLs

@YMourri @AndrewYNg tuning GREAT AI model
<https://deeplearning.ai>

Preprocessing: Handles and URLs

~~@YMourri @AndrewYNg~~ tuning GREAT AI model
~~<https://deeplearning.ai>~~

Preprocessing: Handles and URLs

~~@YMourri @AndrewYNg~~ tuning GREAT AI model
~~https://deeplearning.ai~~



tuning GREAT AI model

Preprocessing: Stemming and lowercasing

tuning GREAT AI model

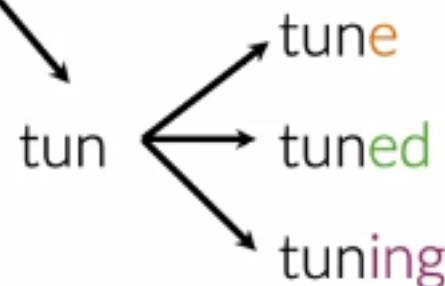
Preprocessing: Stemming and lowercasing

tuning GREAT AI model

tun

Preprocessing: Stemming and lowercasing

tuning GREAT AI model



Preprocessing: Stemming and lowercasing

tuning GREAT AI model

tun →
tune
tuned
tuning

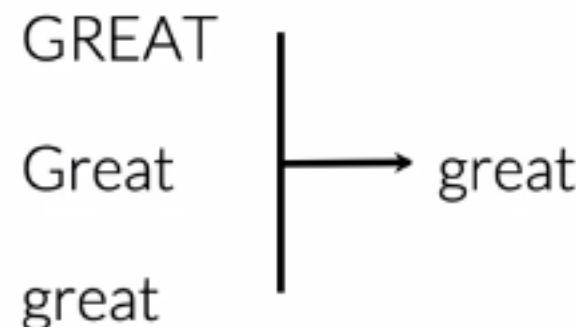
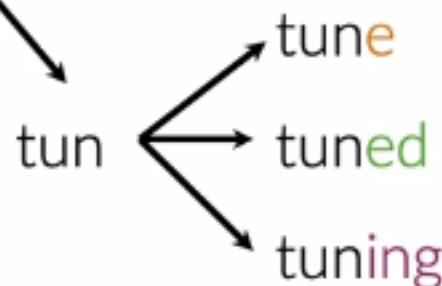
GREAT

Great

great

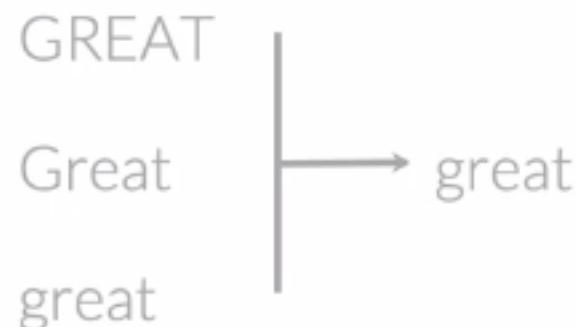
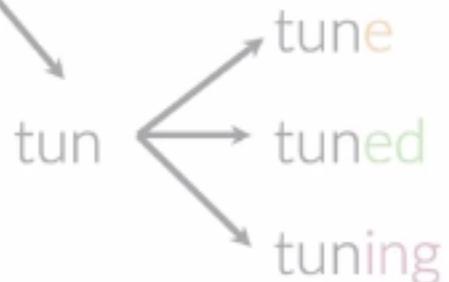
Preprocessing: Stemming and lowercasing

tuning GREAT AI model



Preprocessing: Stemming and lowercasing

tuning GREAT AI model



Preprocessed tweet:
[tun, great, ai, model]

General overview

I am Happy Because i am learning NLP @deeplearning

↓
Preprocessing

[happy, learn, nlp]

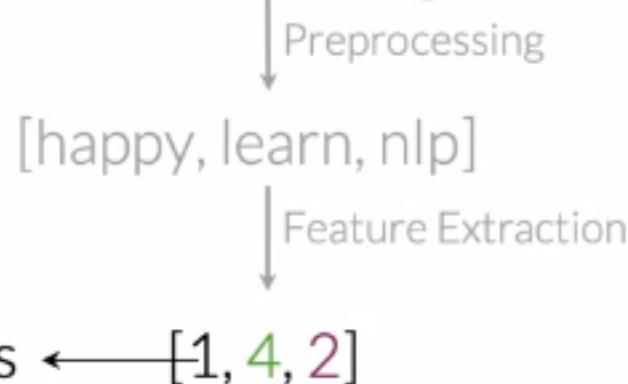
General overview

I am Happy Because i am learning NLP @deeplearning

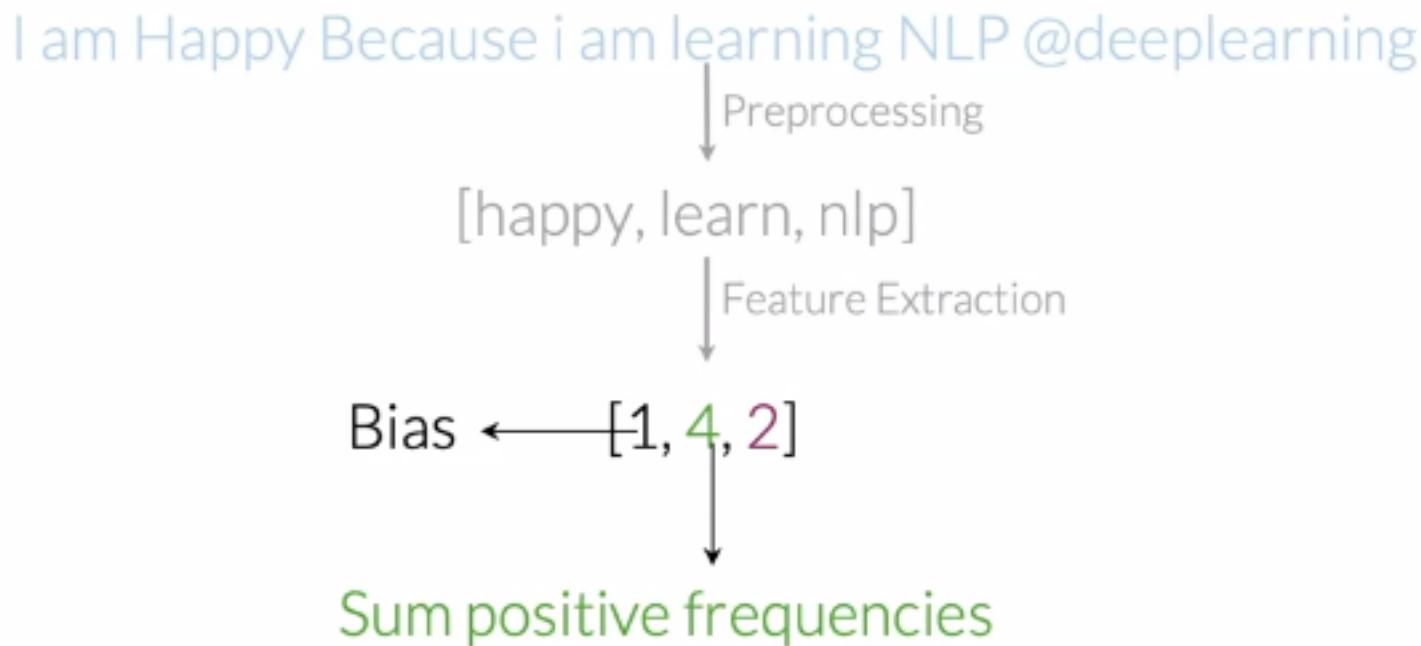


General overview

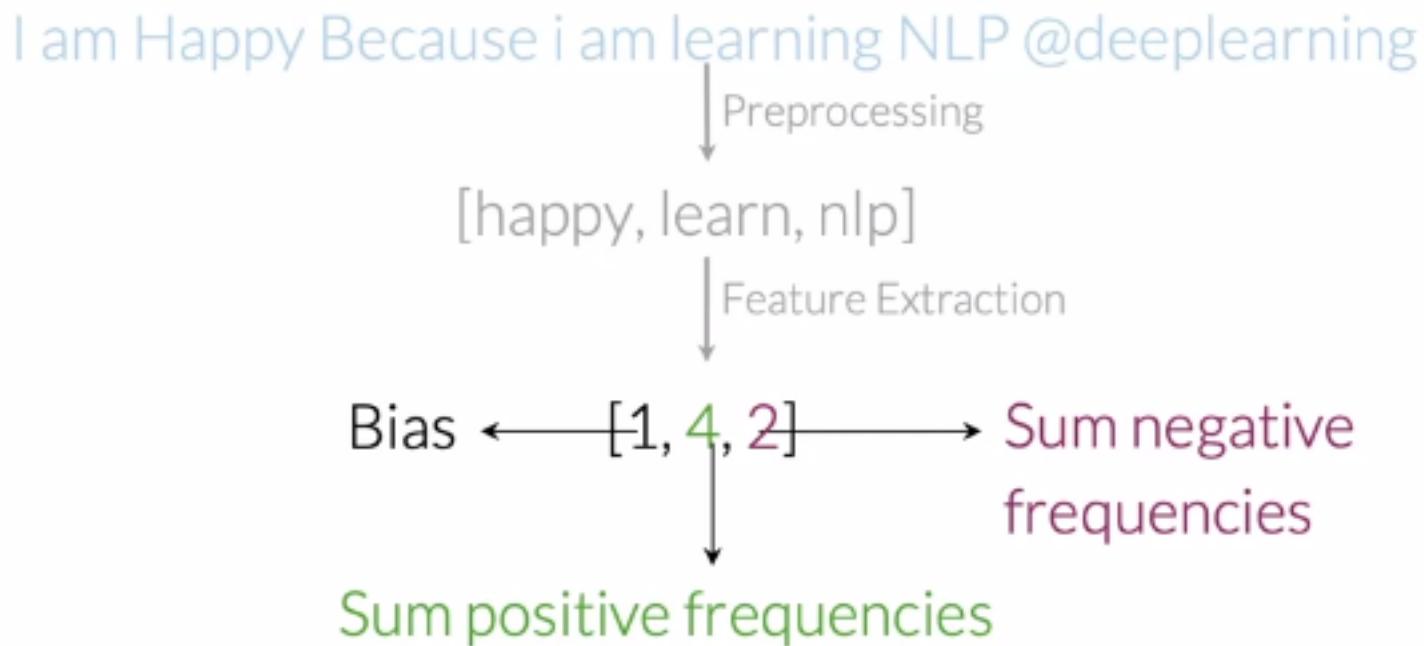
I am Happy Because i am learning NLP @deeplearning



General overview



General overview



General overview

I am Happy Because i am
learning NLP
@deeplearning

I am sad not learning NLP

...

I am sad :(

General overview

I am Happy Because i am
learning NLP
@deeplearning

[happy, learn, nlp]

[sad, not, learn, nlp]

I am sad not learning NLP → ...

[sad]

...
I am sad :(

General overview

I am Happy Because i am
learning NLP
@deeplearning

[happy, learn, nlp]

[[1, 40, 20],
 [1, 20, 50],

I am sad not learning NLP → ...

[sad, not, learn, nlp]

→ ...

[sad]

[1, 5, 35]]

I am sad :(

General overview

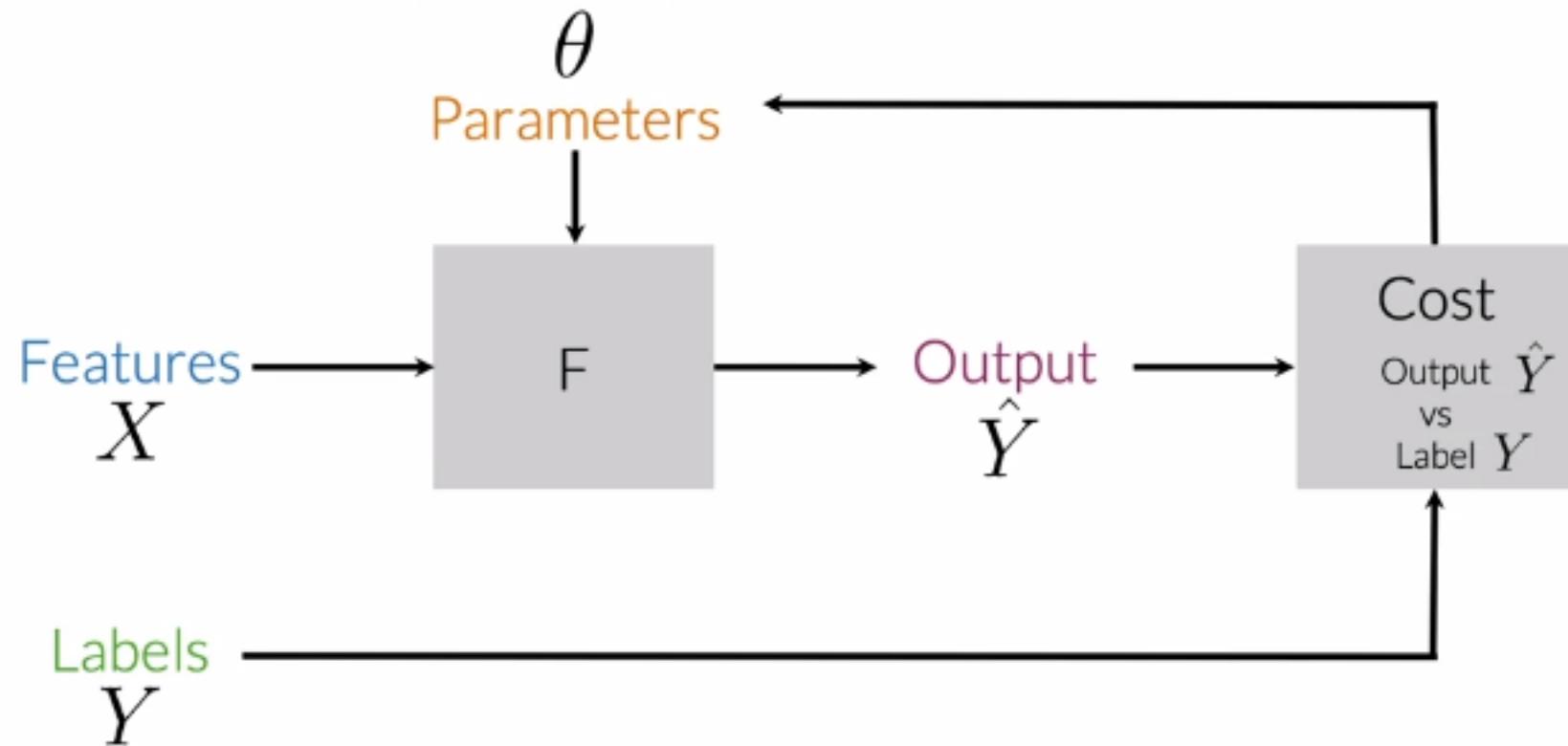
$$\begin{bmatrix} 1 & X_1^{(1)} & X_2^{(1)} \\ 1 & X_1^{(2)} & X_2^{(2)} \\ \vdots & \vdots & \vdots \\ 1 & X_1^{(m)} & X_2^{(m)} \end{bmatrix} \longleftrightarrow$$

[[1, 40, 20],
 [1, 20, 50],
 ...
 [1, 5, 35]]

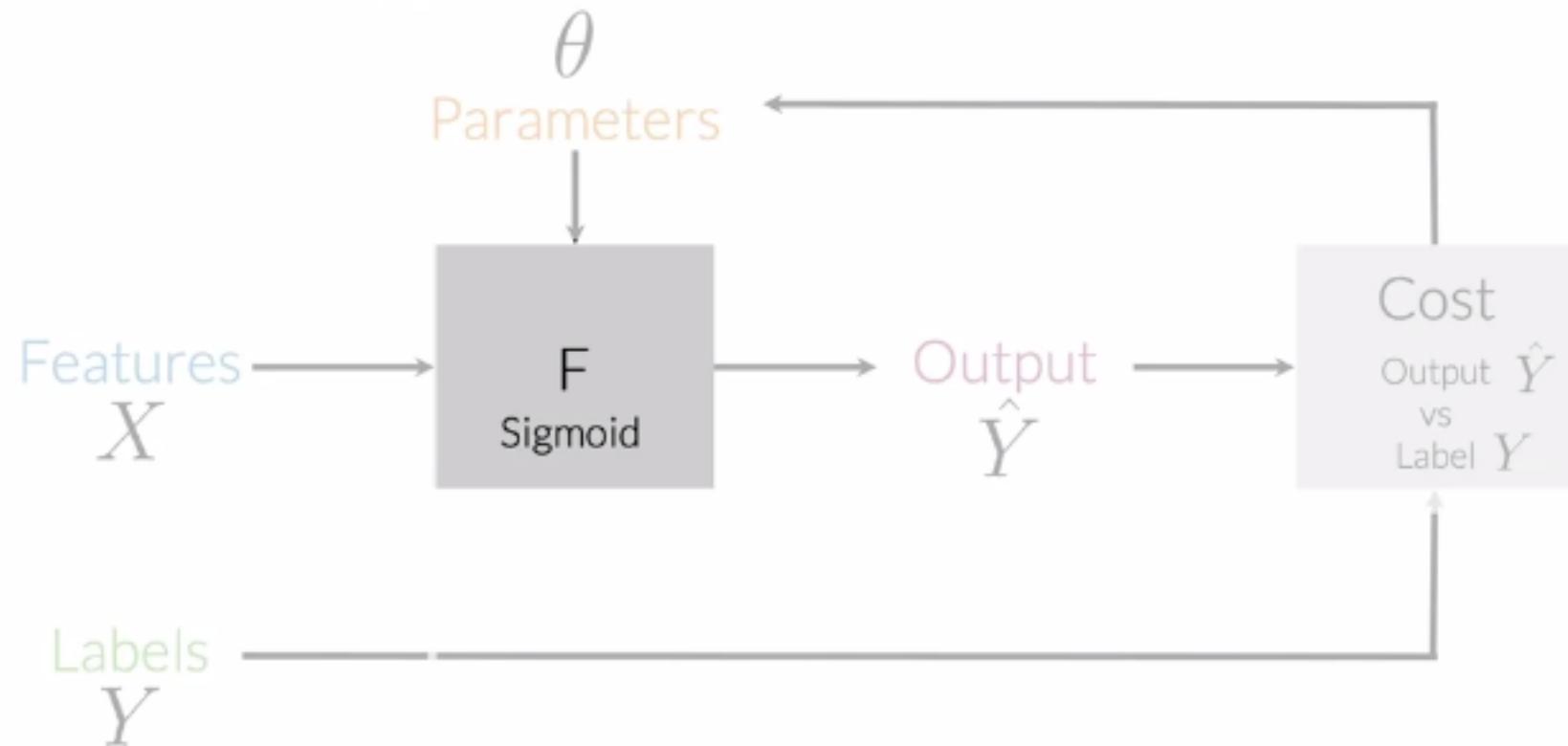
General Implementation

```
freqs = build_freqs(tweets,labels) #Build frequencies dictionary  
X = np.zeros((m,3)) #Initialize matrix X  
  
for i in range(m): #For every tweet  
    p_tweet = process_tweet(tweets[i]) #Process tweet  
    X[i,:] = extract_features(p_tweet,freqs) #Extract Features
```

Overview of logistic regression



Overview of logistic regression

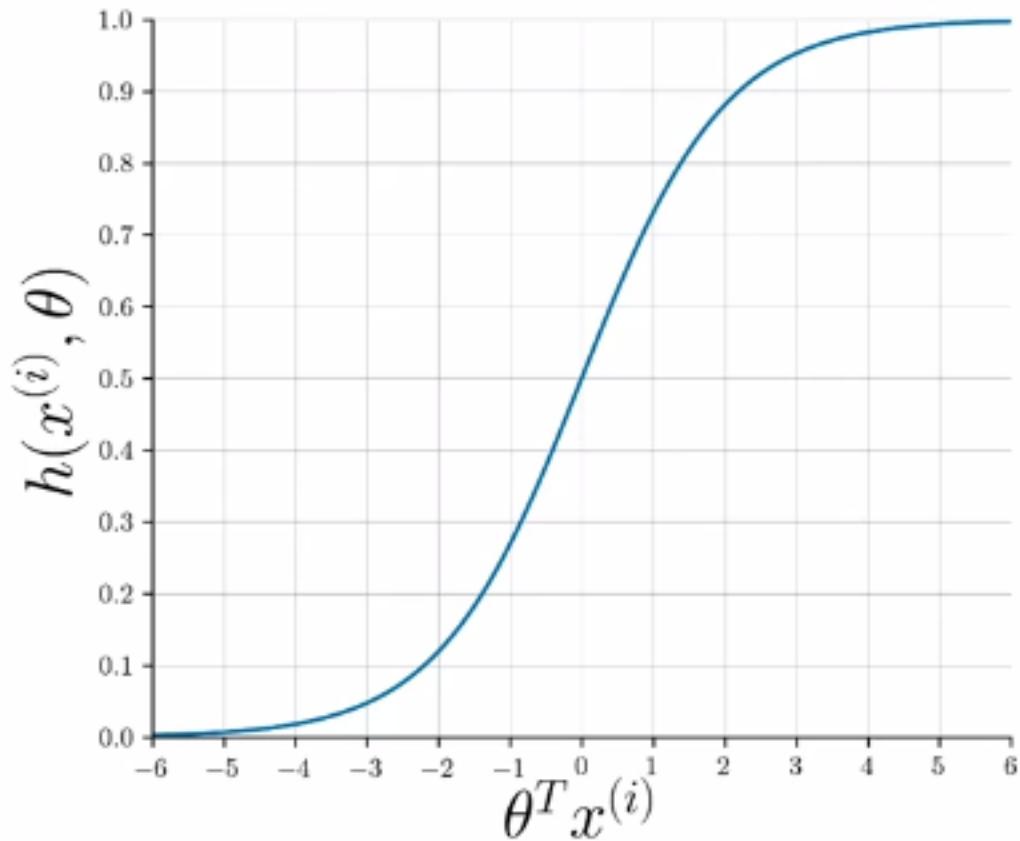


Overview of logistic regression

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$

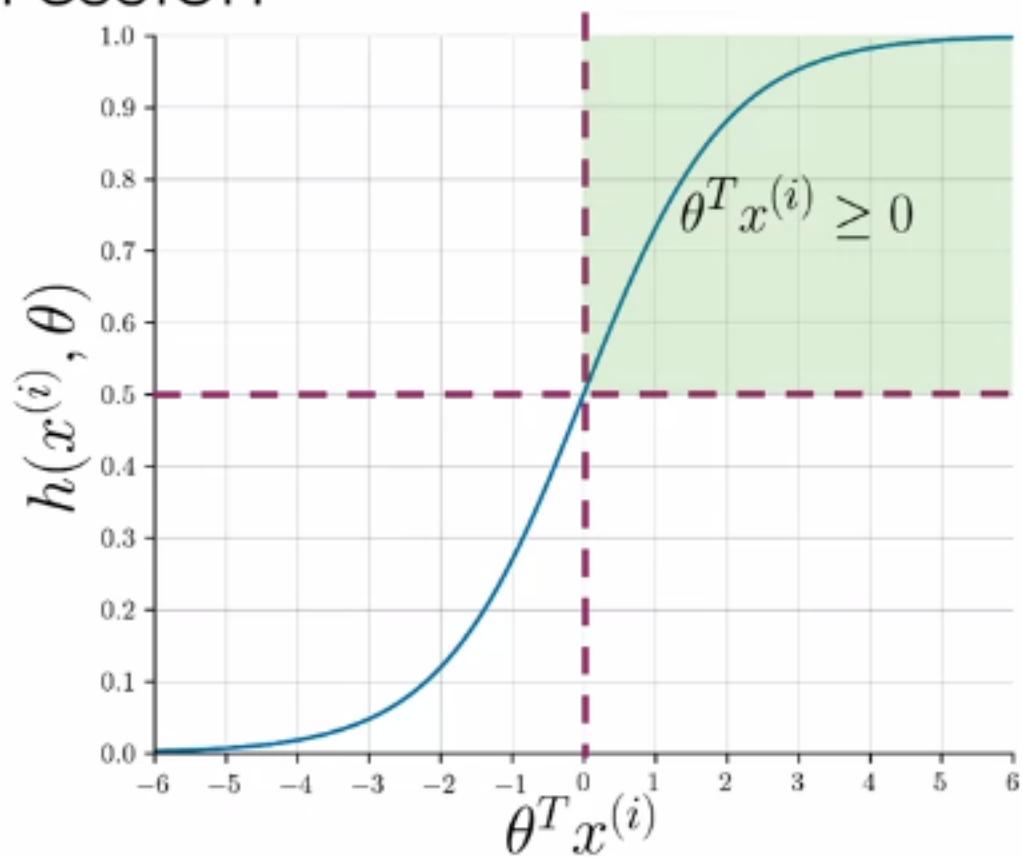
Overview of logistic regression

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$



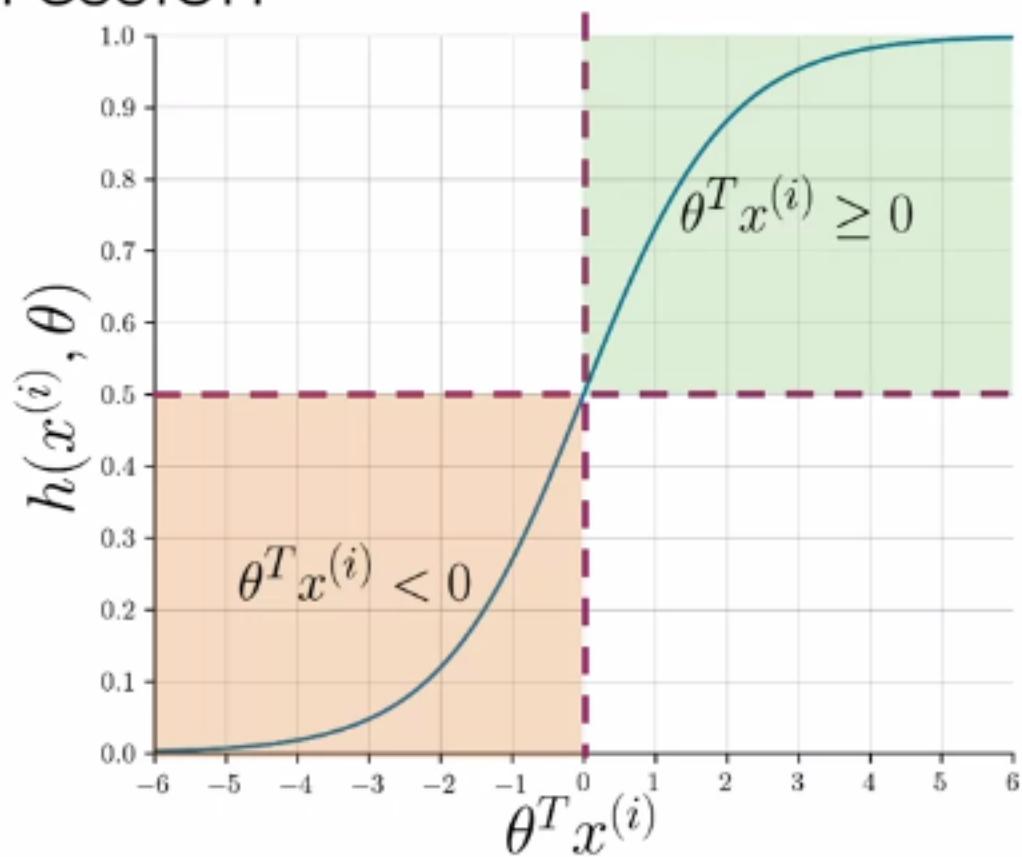
Overview of logistic regression

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$



Overview of logistic regression

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$



Overview of logistic regression

@YMourri and

@AndrewYNg are tuning a
GREAT AI model

Overview of logistic regression

@YMourri and

@AndrewYNg are tuning a

GREAT AI model



[tun, ai, great, model]

Overview of logistic regression

@YMourri and
@AndrewYNg are tuning a
GREAT AI model

[tun, ai, great, model]

$$x^{(i)} = \begin{bmatrix} 1 \\ 3476 \\ 245 \end{bmatrix}$$

Overview of logistic regression

@YMourri and
@AndrewYNg are tuning a
GREAT AI model

[tun, ai, great, model]

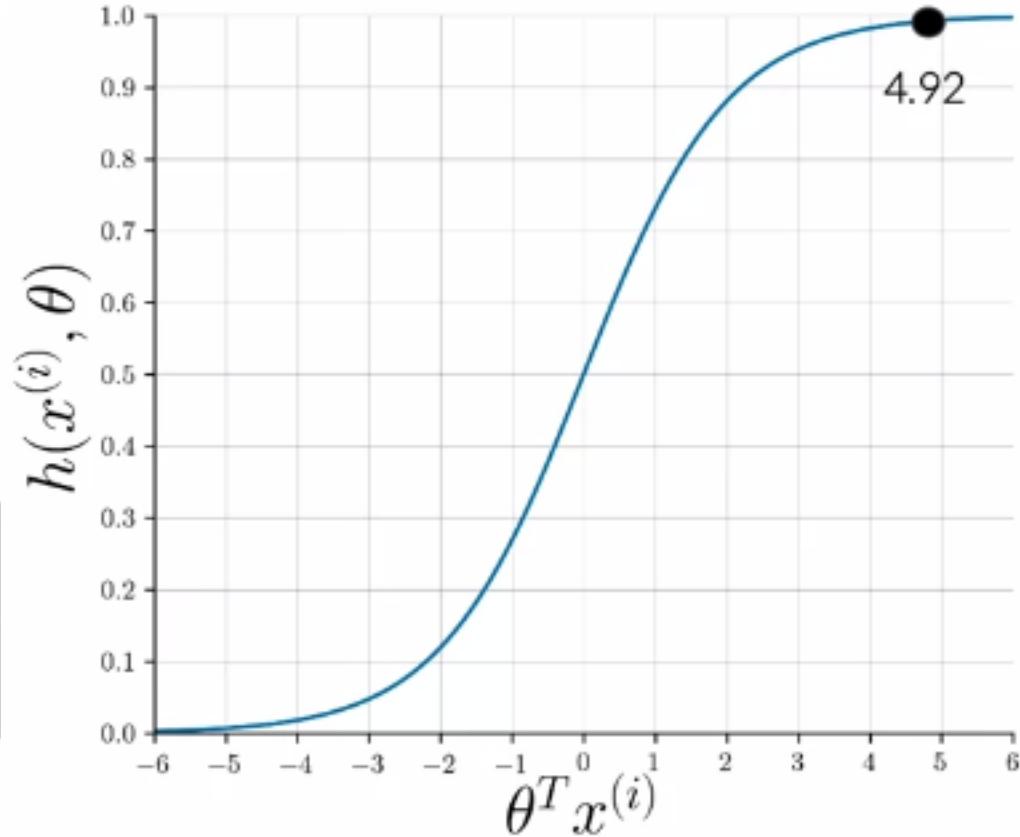
$$x^{(i)} = \begin{bmatrix} 1 \\ 3476 \\ 245 \end{bmatrix} \quad \theta = \begin{bmatrix} 0.00003 \\ 0.00150 \\ -0.00120 \end{bmatrix}$$

Overview of logistic regression

@YMourri and
@AndrewYNg are tuning a
GREAT AI model

[tun, ai, great, model]

$$x^{(i)} = \begin{bmatrix} 1 \\ 3476 \\ 245 \end{bmatrix} \quad \theta = \begin{bmatrix} 0.00003 \\ 0.00150 \\ -0.00120 \end{bmatrix}$$

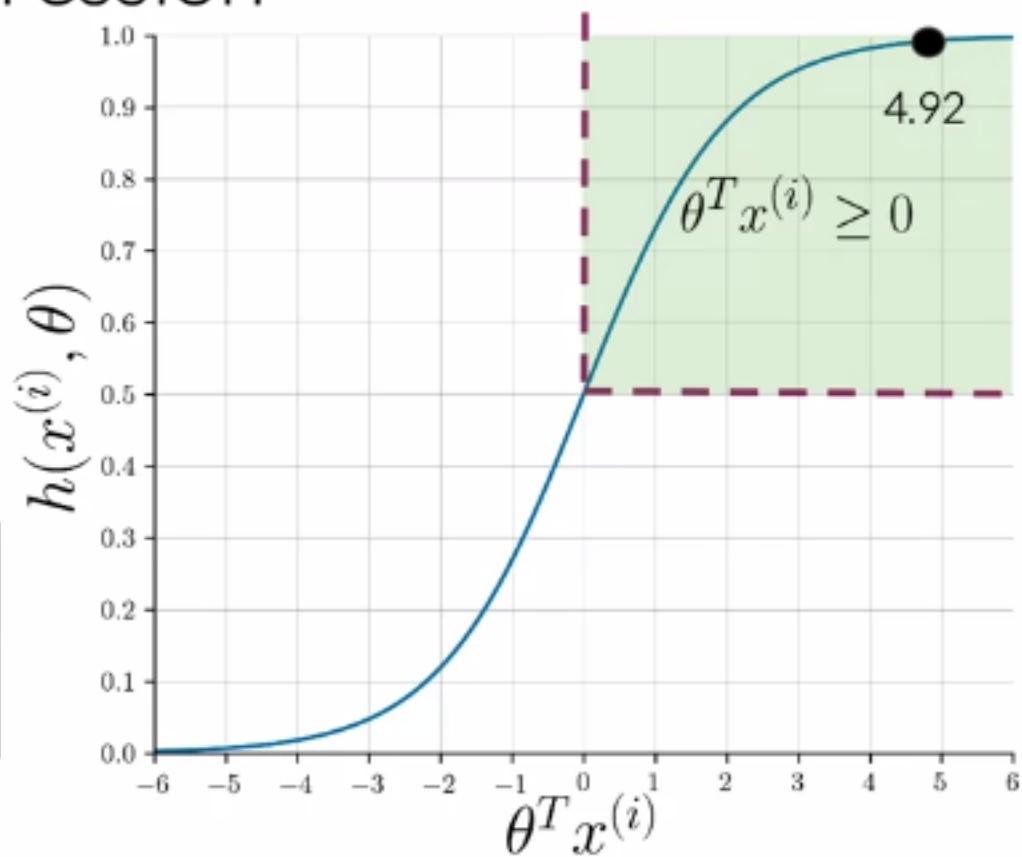


Overview of logistic regression

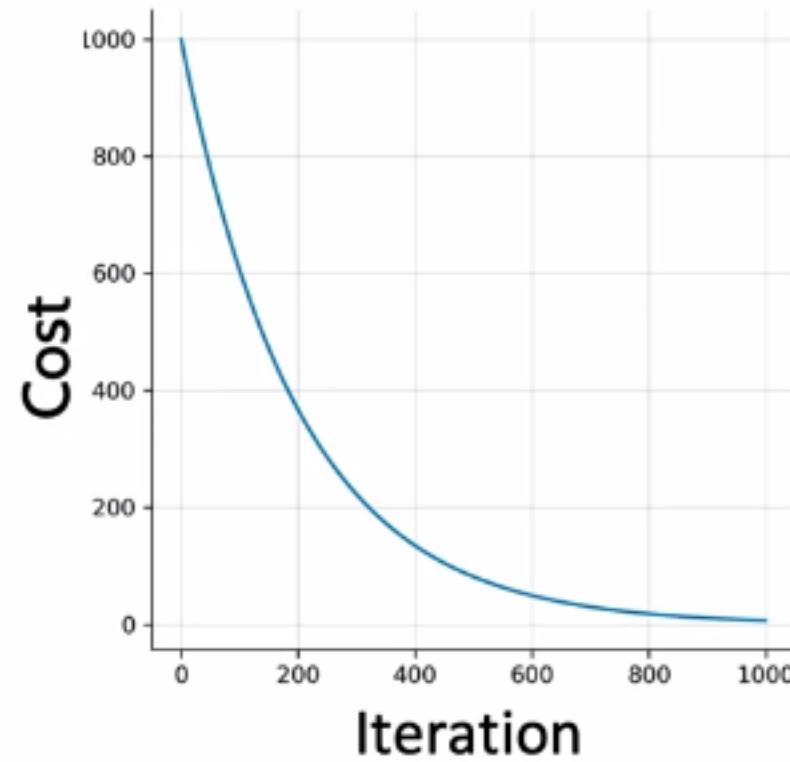
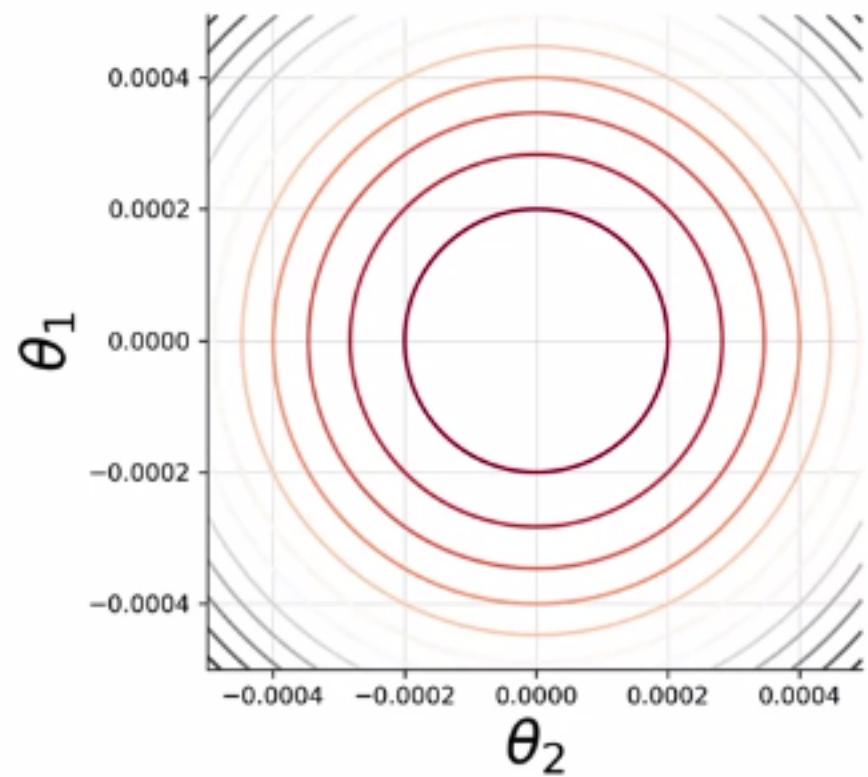
@YMourri and
@AndrewYNg are tuning a
GREAT AI model

[tun, ai, great, model]

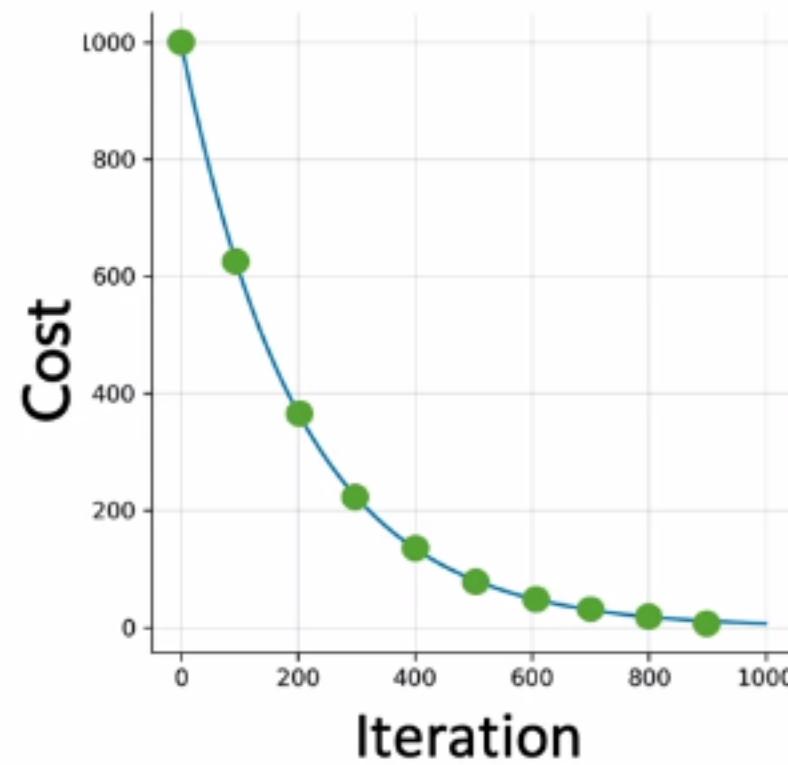
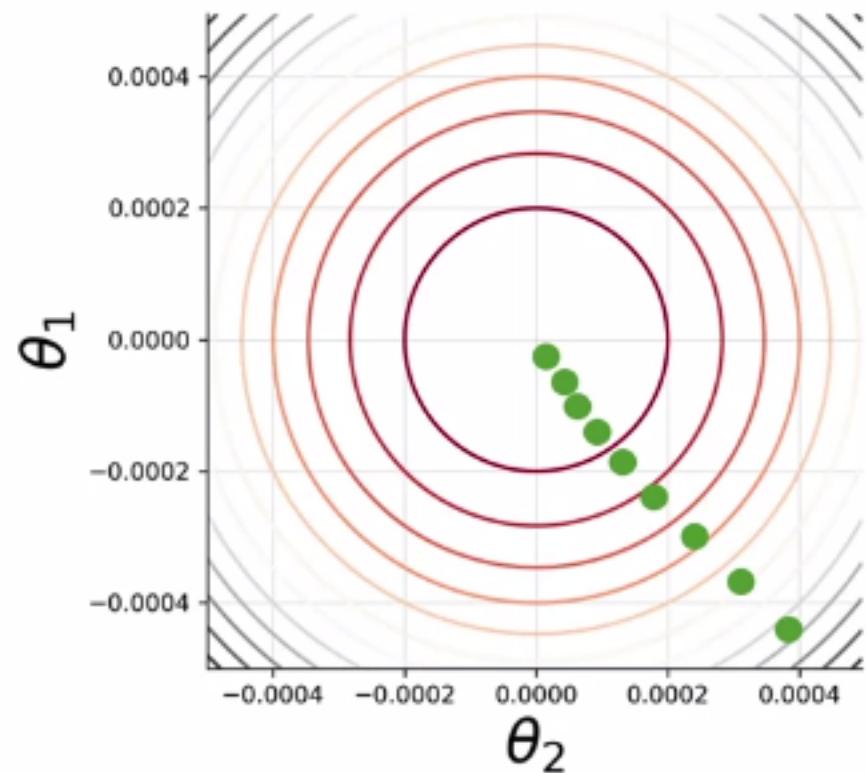
$$x^{(i)} = \begin{bmatrix} 1 \\ 3476 \\ 245 \end{bmatrix} \quad \theta = \begin{bmatrix} 0.00003 \\ 0.00150 \\ -0.00120 \end{bmatrix}$$



Training LR



Training LR



Training LR

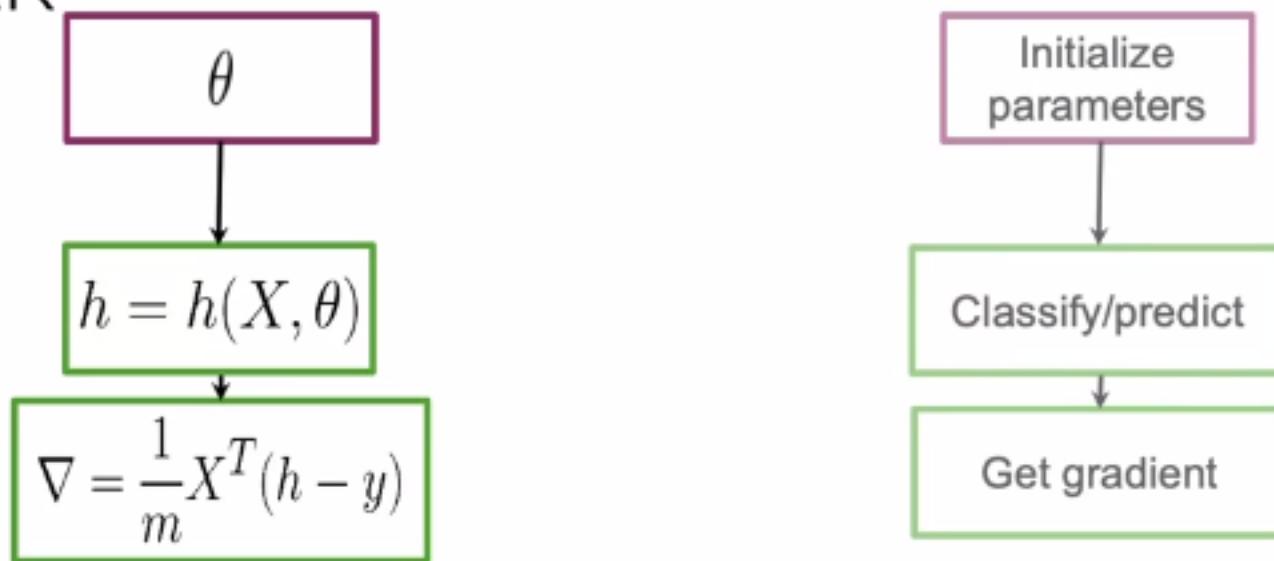
θ

Initialize
parameters

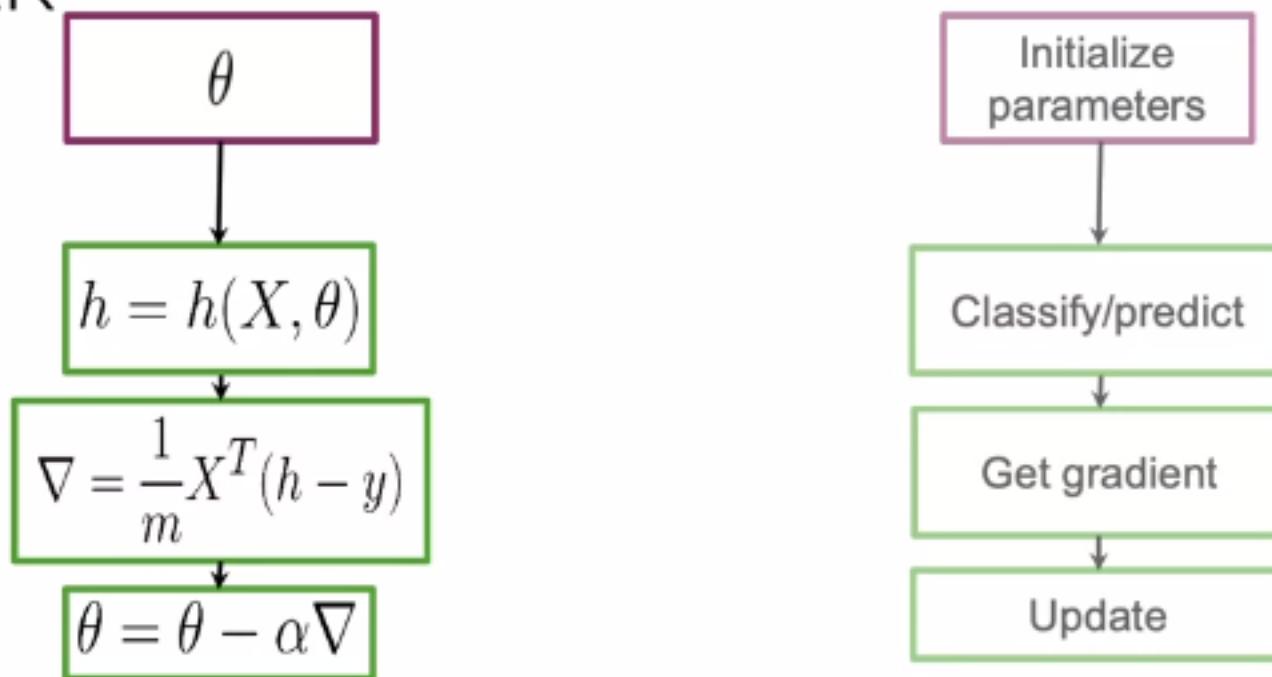
Training LR



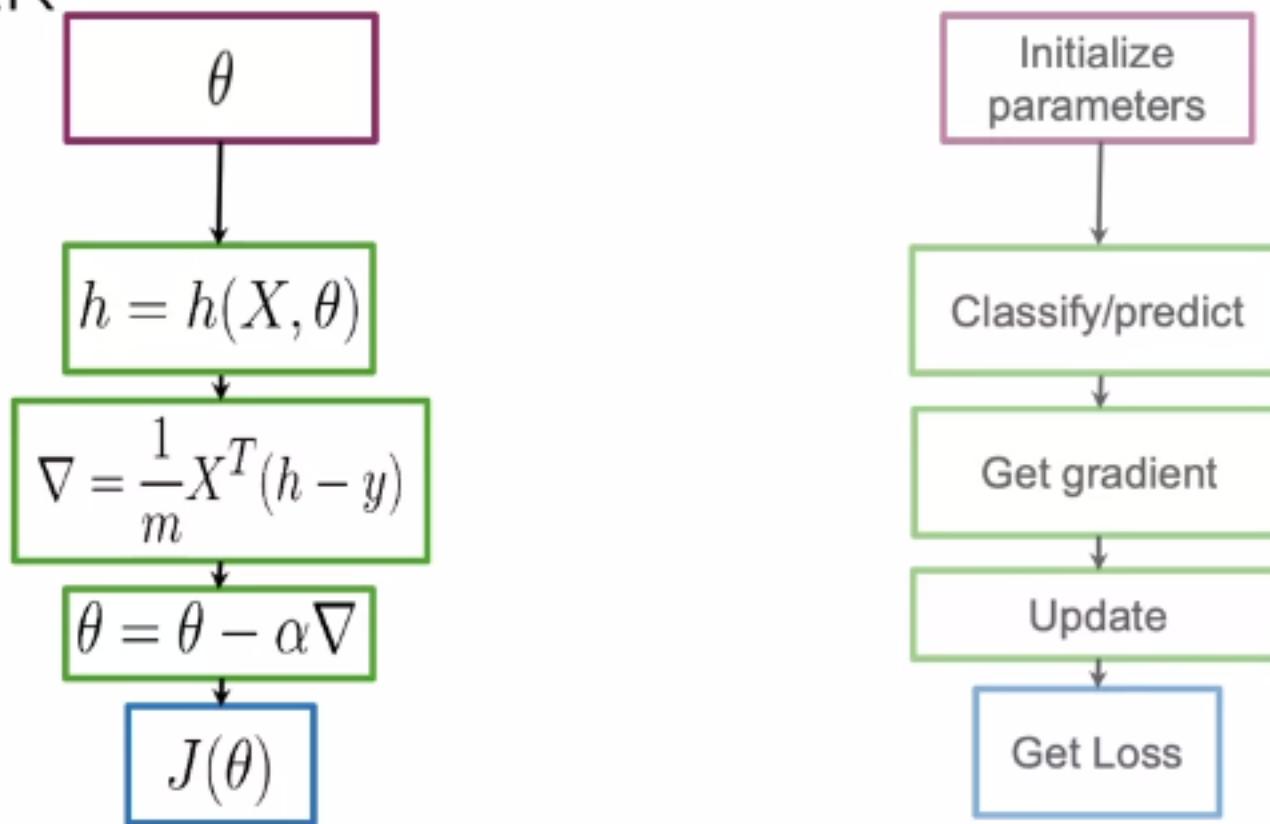
Training LR



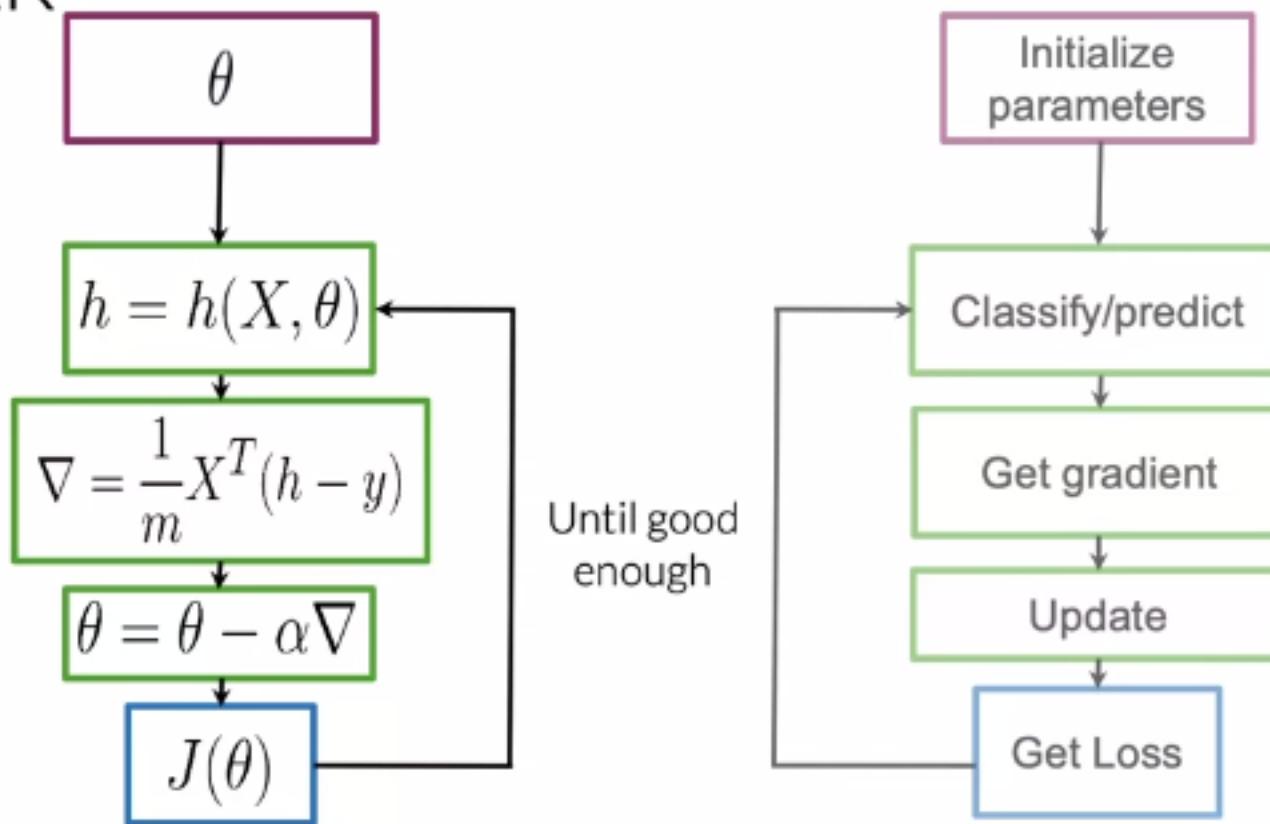
Training LR



Training LR



Training LR



Testing logistic regression

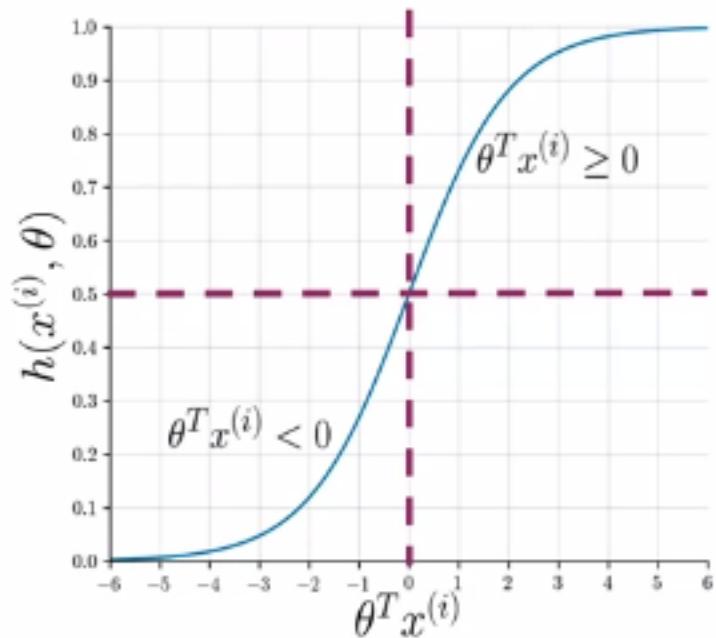
- X_{val} Y_{val} θ

Testing logistic regression

- X_{val} Y_{val} θ
 $h(X_{val}, \theta)$

Testing logistic regression

- $X_{val} \ Y_{val} \ \theta$

$$h(X_{val}, \theta)$$
$$pred = h(X_{val}, \theta) \geq 0.5$$


Testing logistic regression

- X_{val} Y_{val} θ

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\begin{bmatrix} 0.3 \\ 0.8 \\ 0.5 \\ \vdots \\ h_m \end{bmatrix}$$

Testing logistic regression

- X_{val} Y_{val} θ

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\begin{bmatrix} 0.3 \\ 0.8 \\ 0.5 \\ \vdots \\ h_m \end{bmatrix} \geq 0.5$$

Testing logistic regression

- $X_{val} \ Y_{val} \ \theta$

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\begin{bmatrix} 0.3 \\ 0.8 \\ 0.5 \\ \vdots \\ h_m \end{bmatrix} \geq 0.5 = \begin{bmatrix} 0.3 \geq 0.5 \\ 0.8 \geq 0.5 \\ 0.5 \geq 0.5 \\ \vdots \\ pred_m \geq 0.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ pred_m \end{bmatrix}$$

Testing logistic regression

- $X_{val} \ Y_{val} \ \theta$

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\sum_{i=1}^m \frac{(pred^{(i)} == y_{val}^{(i)})}{m}$$

Testing logistic regression

- $X_{val} \ Y_{val} \ \theta$

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\sum_{i=1}^m \frac{(pred^{(i)} == y_{val}^{(i)})}{m}$$

$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ pred_m \end{bmatrix} == \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ Y_{val_m} \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ \vdots \\ pred_m == Y_{val_m} \end{bmatrix}$$

Testing logistic regression

- $X_{val} \ Y_{val} \ \theta$

$$h(X_{val}, \theta)$$

$$pred = h(X_{val}, \theta) \geq 0.5$$

$$\sum_{i=1}^m \frac{(pred^{(i)} == y_{val}^{(i)})}{m}$$

Testing logistic regression

$$Y_{val} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad pred = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Testing logistic regression

$$Y_{val} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad pred = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (Y_{val} == pred) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Testing logistic regression

$$Y_{val} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad pred = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (Y_{val} == pred) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{accuracy} = \frac{4}{5} = 0.8$$

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

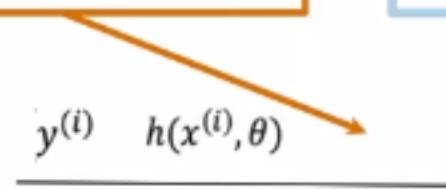
Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

Cost function for logistic regression

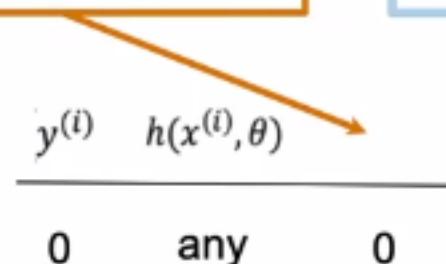
$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$


The diagram illustrates the cost function for logistic regression. It shows the formula $J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$. Two terms in the sum are highlighted: $y^{(i)} \log h(x^{(i)}, \theta)$ is enclosed in an orange box, and $(1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))$ is enclosed in a blue box. Orange and blue arrows point from these highlighted terms to a single horizontal line below them, indicating they are being summed.

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$


The diagram illustrates the components of the cost function. It shows a red arrow pointing from the term $y^{(i)} \log h(x^{(i)}, \theta)$ to a horizontal line above three boxes labeled 0, any, and 0. This indicates that the term is zero when $y^{(i)} = 0$ and $h(x^{(i)}, \theta) = 1$, and is "any" (non-zero) otherwise.

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

| $y^{(i)}$ | $h(x^{(i)}, \theta)$ | |
|-----------|----------------------|----|
| 0 | any | 0 |
| 1 | 0.99 | ~0 |

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

| $y^{(i)}$ | $h(x^{(i)}, \theta)$ | |
|-----------|----------------------|------|
| 0 | any | 0 |
| 1 | 0.99 | ~0 |
| 1 | ~0 | -inf |

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

$y^{(i)} - h(x^{(i)}, \theta)$



Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

$$\frac{y^{(i)} - h(x^{(i)}, \theta)}{1 \quad \text{any} \quad 0}$$


Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

| $y^{(i)}$ | $h(x^{(i)}, \theta)$ | |
|-----------|----------------------|----|
| 1 | any | 0 |
| 0 | 0.01 | ~0 |

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

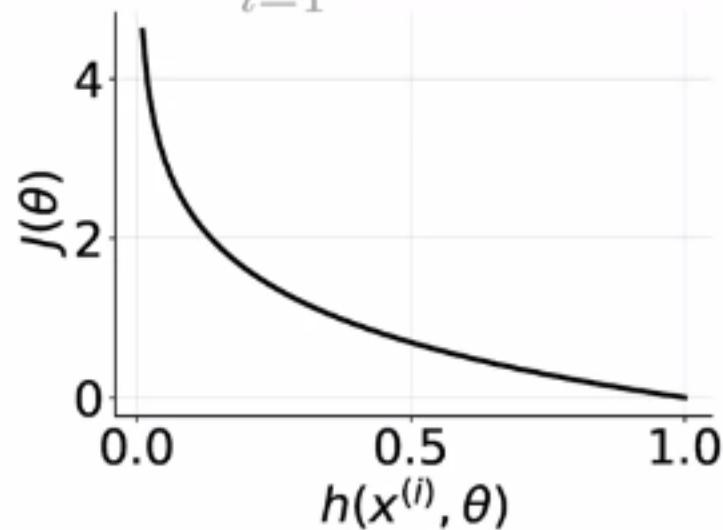
| $y^{(i)}$ | $h(x^{(i)}, \theta)$ | |
|-----------|----------------------|------|
| 1 | any | 0 |
| 0 | 0.01 | ~0 |
| 0 | ~1 | -Inf |

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$



Cost function for logistic regression

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log h(x^{(i)}, \theta) + (1 - y^{(i)}) \log(1 - h(x^{(i)}, \theta))]$$

