Instituto de Computação
UNIVERSIDADE ESTADUAL DE CAMPINAS

Capacitação profissional em tecnologias de Inteligência Artificial

Machine Learning Overview

Prof. Edson Borin

https://www.ic.unicamp.br/~edson

Institute of Computing - UNICAMP

Common ML Algorithms

- **Machine learning**
  - **Supervised Learning**
    - Classification
      - Logistic regression
      - SVM
      - Neural Network
      - Decision Tree
      - Random Forest
      - GBDT
      - KNN
      - Naive Bayes
    - Regression
      - Linear regression
      - SVM
      - Neural Network
      - Decision Tree
      - Random Forest
      - GBDT
  - **Unsupervised Learning**
    - Clustering
      - K-means
      - Hierarchical clustering
      - Density-based clustering
    - Others
      - Correlation rule
      - Principal component analysis (PCA)
      - Gaussian mixture models (GMM)

Logistic Regression

Supervised Learning Algorithms
Logistic Regression

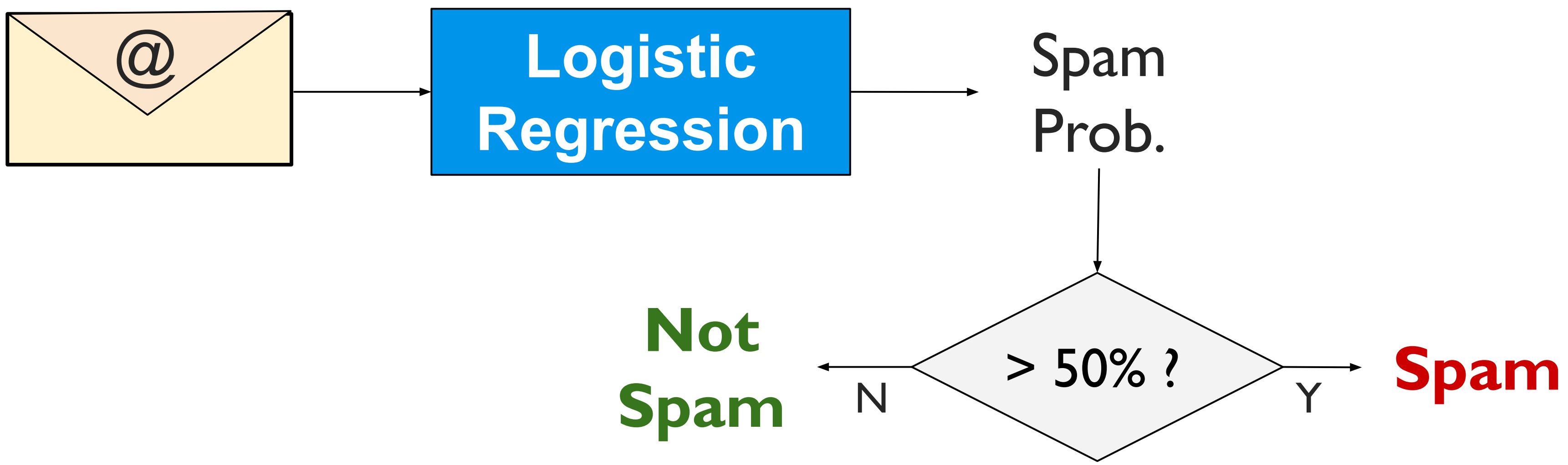
Regression vs Classification recap

Regression	Classification
• The ML model predicts the output for the given input • $f: R^n \rightarrow R$ • Ex: Estimate insurance premium or house price	• The ML model specify which of the k categories the given input belongs to • $f: R^n \rightarrow (1, 2, \ldots, k)$ • Ex: Classify pictures

Supervised Learning Algorithms
Logistic Regression

Logistic regression is used to solve classification problems

- Ex: Classify whether an email is spam or not

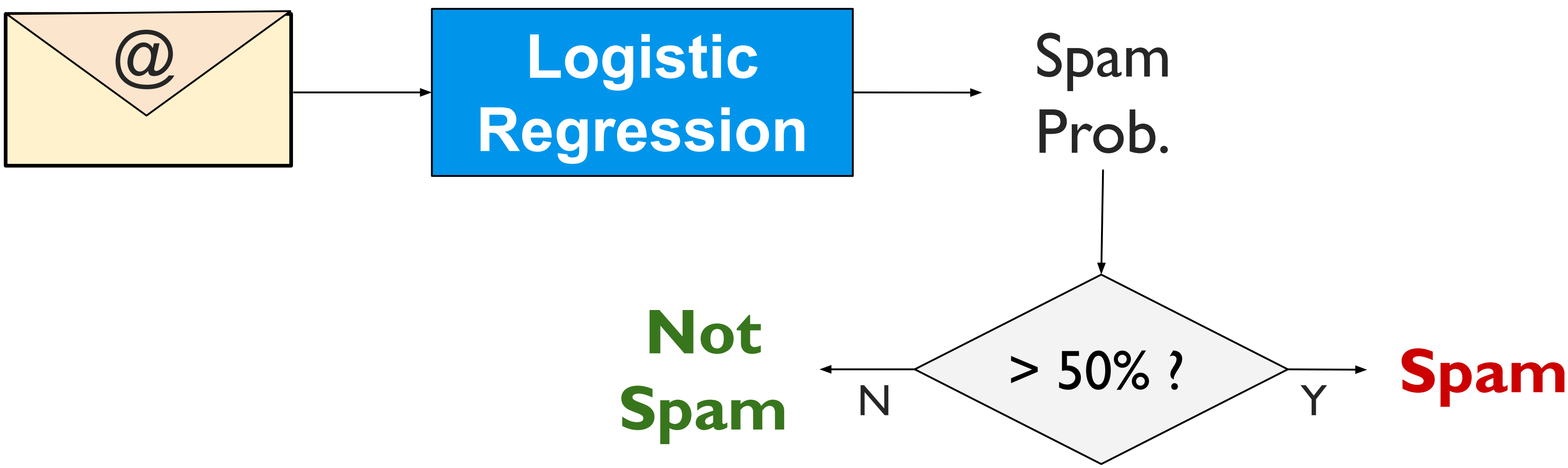


Supervised Learning Algorithms
Logistic Regression

Logistic regression is used to solve c
problems

- Ex: Classify whether an email is s

Logistic Regression works by estimating probabilities!

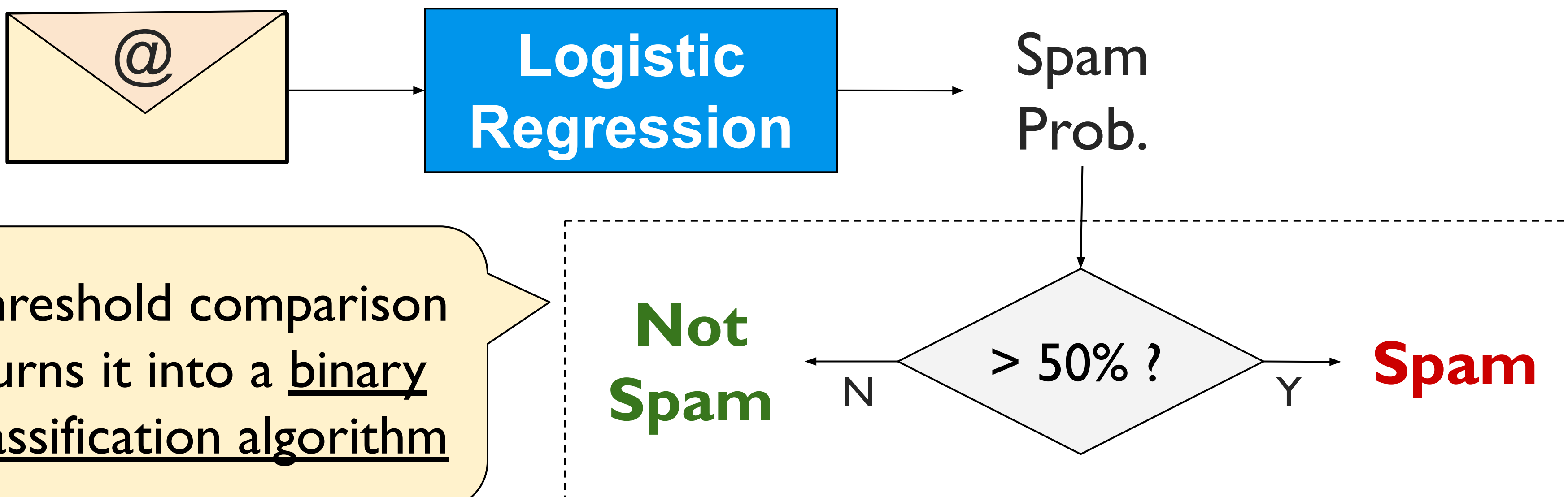


Supervised Learning Algorithms
Logistic Regression

Logistic regression is used to solve c
problems

- Ex: Classify whether an email is s

Logistic Regression works by estimating probabilities!



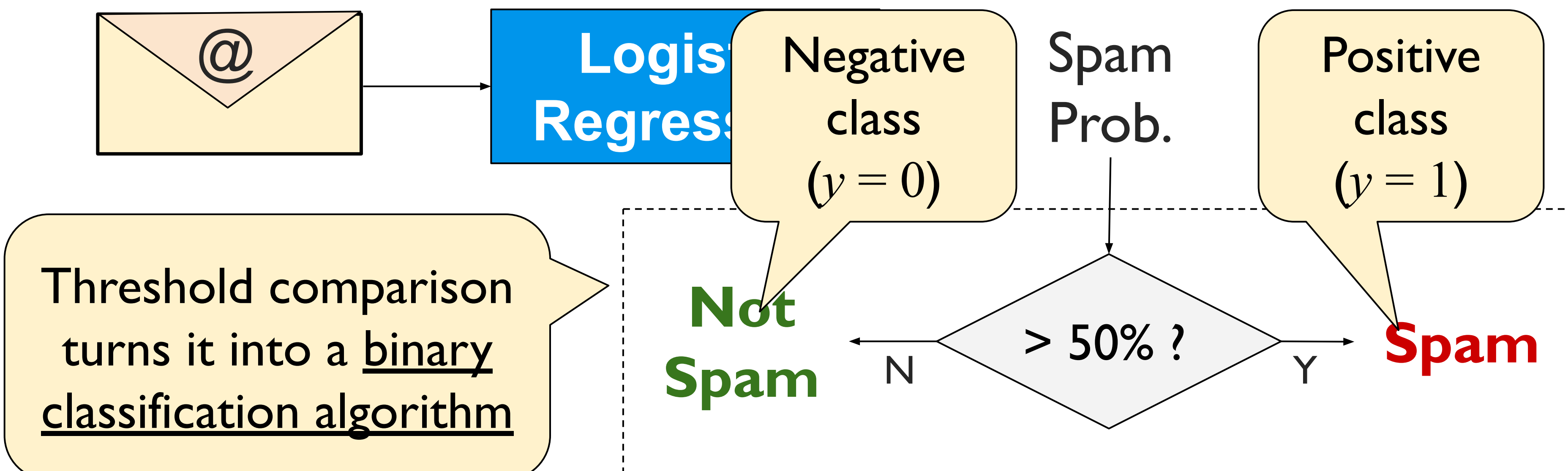
Threshold comparison turns it into a <u>binary classification algorithm</u>

Supervised Learning Algorithms
Logistic Regression

Logistic regression is used to solve c
problems

- Ex: Classify whether an email is s

Logistic Regression works by estimating probabilities!

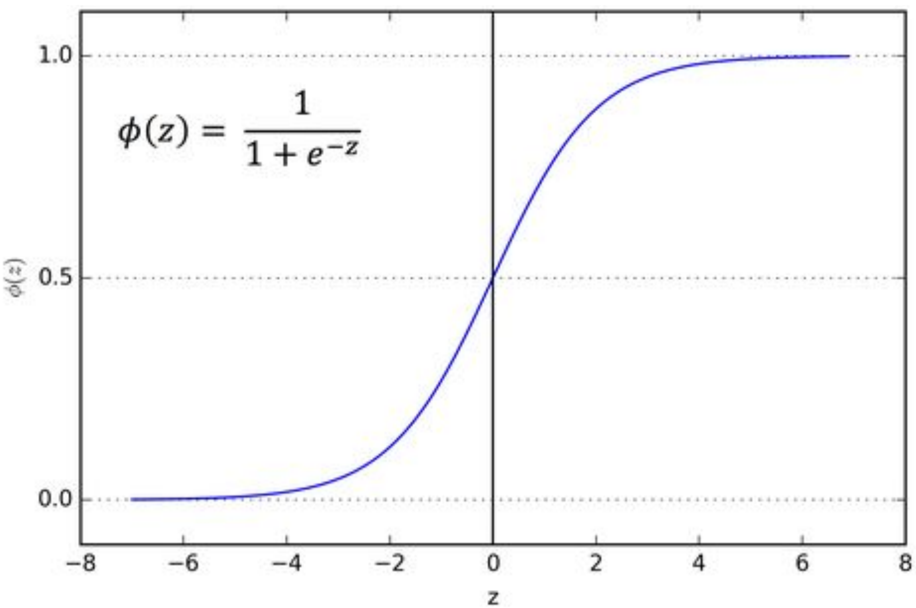


Logistic regression model

- Log. R. model = linear regression model combined with the sigmoid function!
  - Linear regression model: $\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n$

Logistic regression model

- Log. R. model = linear regression model combined with the sigmoid function!
  - Linear regression model: $\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n$
  - Sigmoid function: $\sigma(z) = 1 / (1+e^{-z})$



Logistic regression model

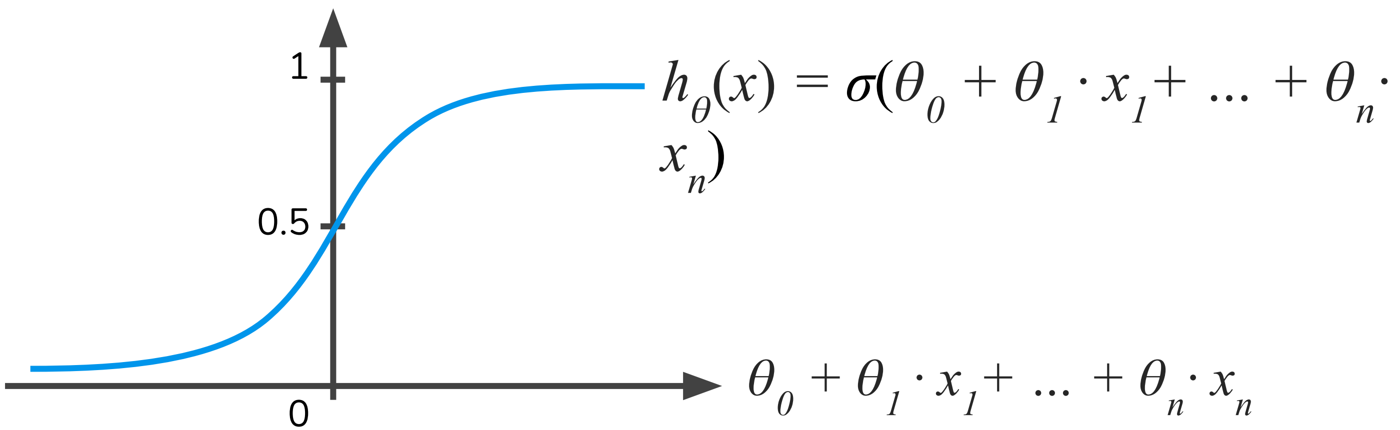
- Log. R. model = linear regression model combined with the sigmoid function!
  - Linear regression model: $\hat{y} = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n$
  - Sigmoid function: $\sigma(z) = 1 / (1+e^{-z})$

- Logistic regression model: $\hat{y}=\sigma(\theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n)$

$$\hat{y} = h_\theta(x) = \frac{1}{1+e^{-(\theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n)}}$$

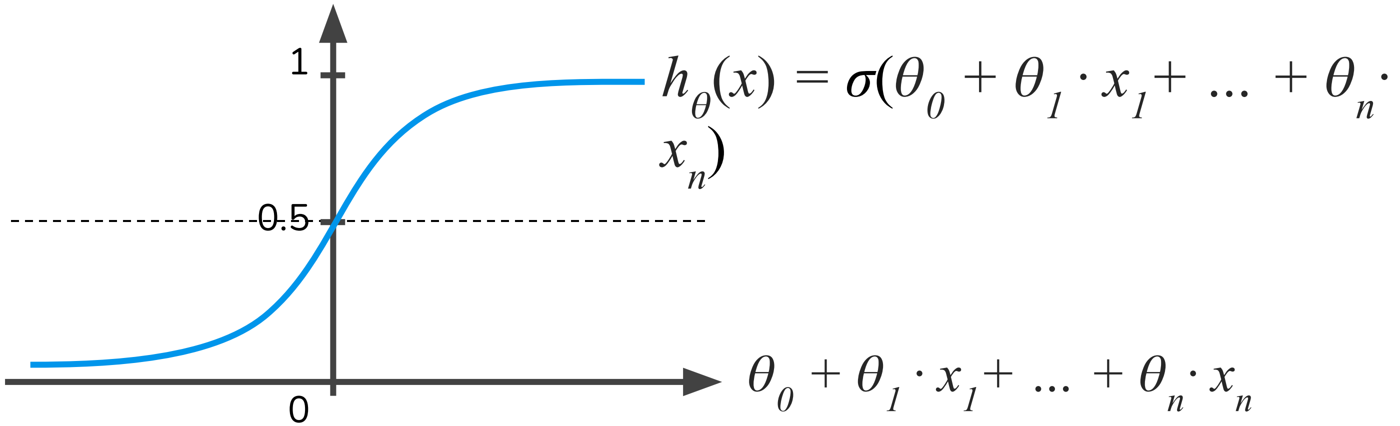
Logistic regression model

- Examples



Logistic regression model

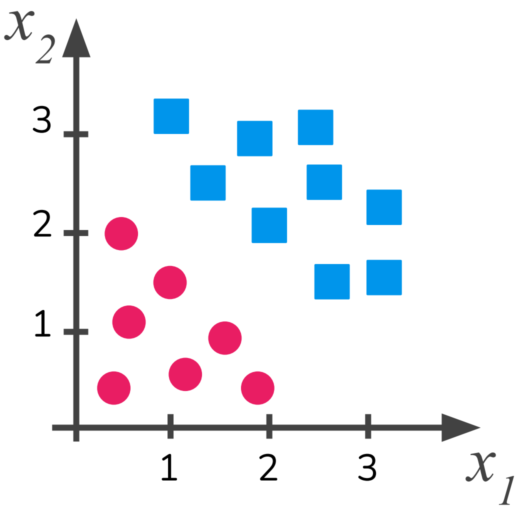
- Examples



In our examples we will use threshold = 0.5
Consequently, if $\theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_n \cdot x_n > 0.0$, then x belongs to the positive class, otherwise, it belongs to the negative class.

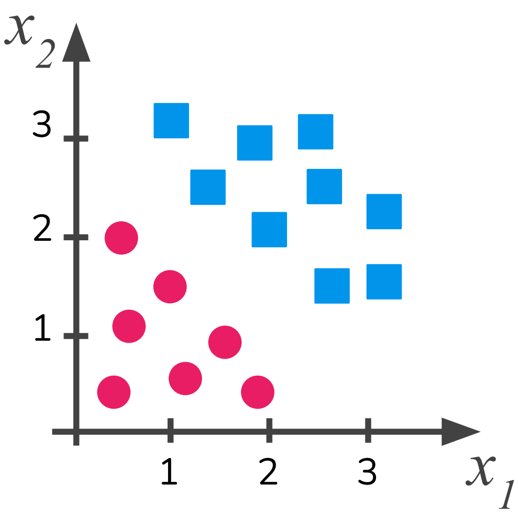
Logistic regression model

- Examples



Logistic regression model

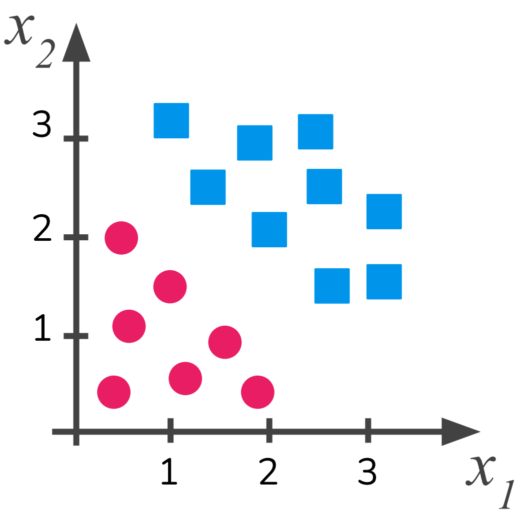
- Examples



$$h_\theta(x) = \sigma(\theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2)$$

Logistic regression model

- Examples

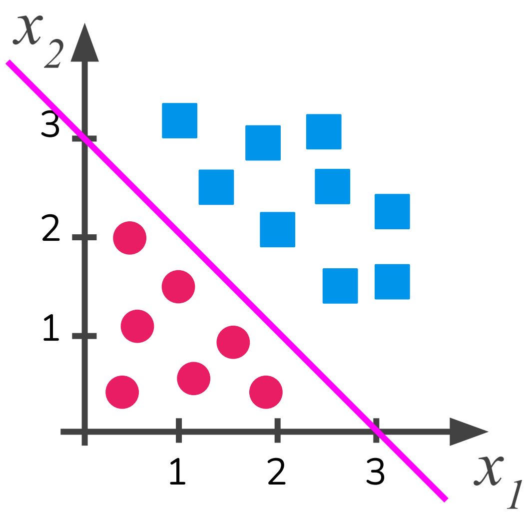


$$h_\theta(x) = \sigma(\overset{-3}{\theta_0} + \overset{1}{\theta_1} \cdot x_1 + \overset{1}{\theta_2} \cdot x_2)$$

Predict "y = 1" if $-3+x_1+x_2 \geq 0$

Logistic regression model

- Examples



$$\begin{array}{c} -3 \quad\quad 1 \quad\quad\quad 1 \\ \Uparrow \quad\quad \Uparrow \quad\quad\quad \Uparrow \end{array}$$

$$h_\theta(x) = \sigma(\theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2)$$

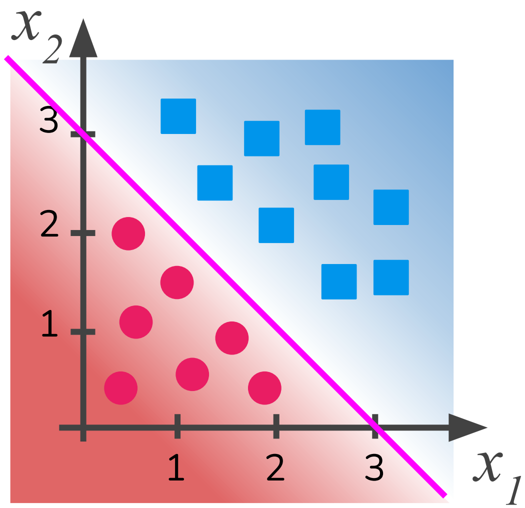
Predict "y = 1" if $\underbrace{-3 + x_1 + x_2 \geq 0}$

Decision boundary

$$x_1 + x_2 = 3$$

Logistic regression model

- Examples



$$h_\theta(x) = \sigma(\underset{-3}{\theta_0} + \underset{1}{\theta_1} \cdot x_1 + \underset{1}{\theta_2} \cdot x_2)$$

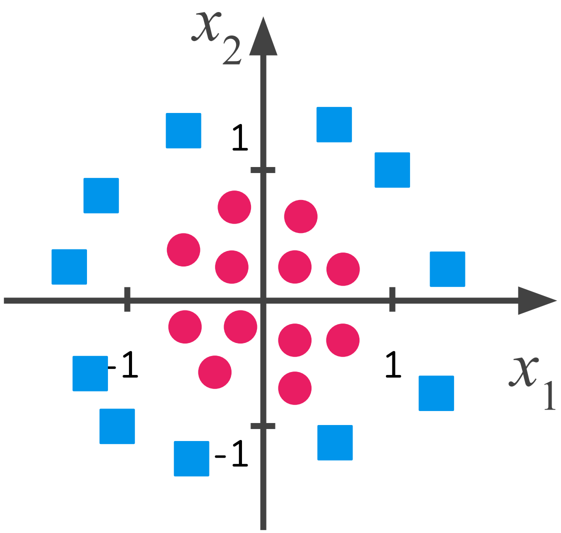
Predict "y = 1" if $\underbrace{-3 + x_1 + x_2 \geq 0}$

Decision boundary

$$x_1 + x_2 = 3$$

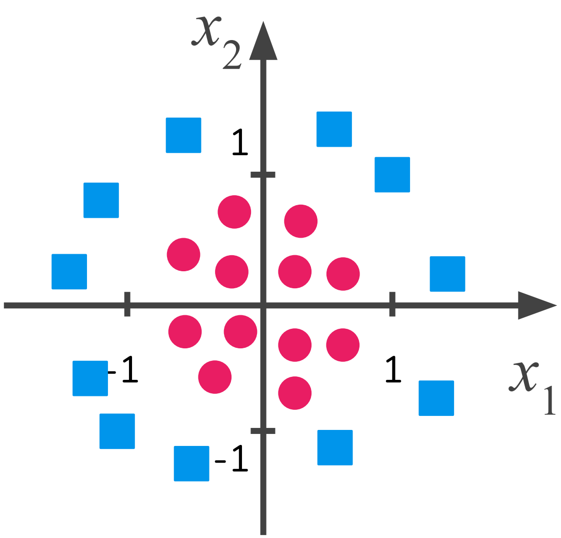
Logistic regression model

- Non-linear Decision Boundaries



Logistic regression model

- Non-linear Decision Boundaries

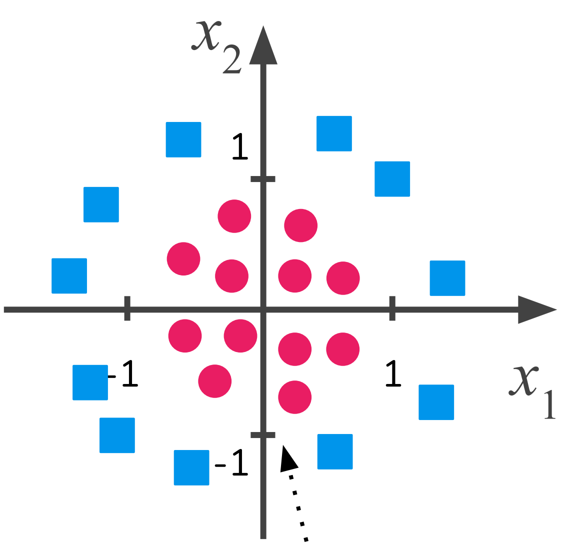


$$h_\theta(x) = \sigma(\theta_0+\theta_1 x_1+\theta_2 x_2+\theta_3 x_1^2+\theta_4 x_2^2)$$

Supervised Learning Algorithms
Logistic Regression

Logistic regression model

- Non-linear Decision Boundaries



$$h_\theta(x) = \sigma(\overset{-1}{\theta_0} + \overset{0}{\theta_1} x_1 + \overset{0}{\theta_2} x_2 + \overset{1}{\theta_3} x_1^2 + \overset{1}{\theta_4} x_2^2)$$

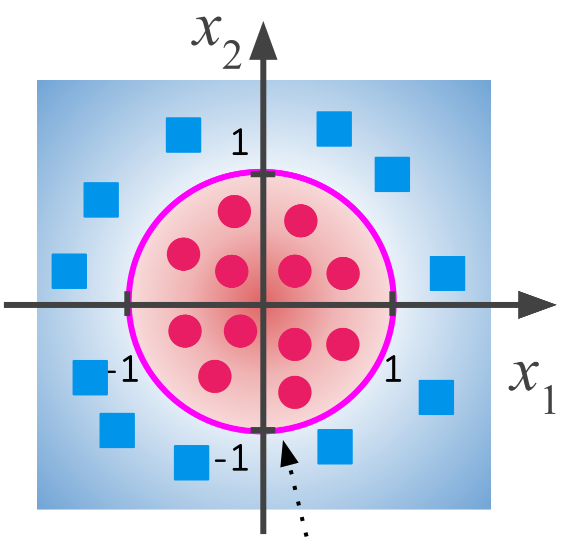
Predict "y = 1" if $\underbrace{-1 + x_1^2 + x_2^2 \geq 0}$

Decision boundary

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

Logistic regression model

- Non-linear Decision Boundaries



$$h_\theta(x) = \sigma(\underset{-1}{\theta_0} + \underset{0}{\theta_1} x_1 + \underset{0}{\theta_2} x_2 + \underset{1}{\theta_3} x_1^2 + \underset{1}{\theta_4} x_2^2)$$

Predict "y = 1" if $\underbrace{-1 + x_1^2 + x_2^2}_{} \geq 0$

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

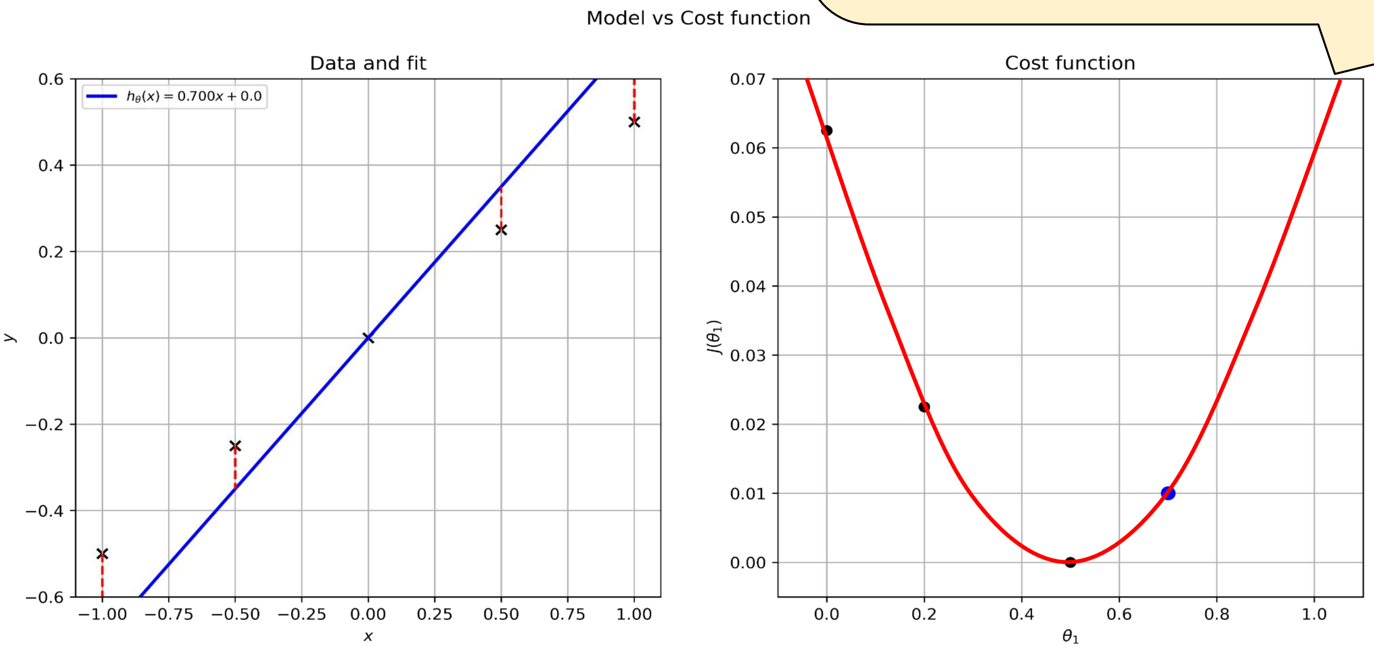
Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$

Training a logistic regression model

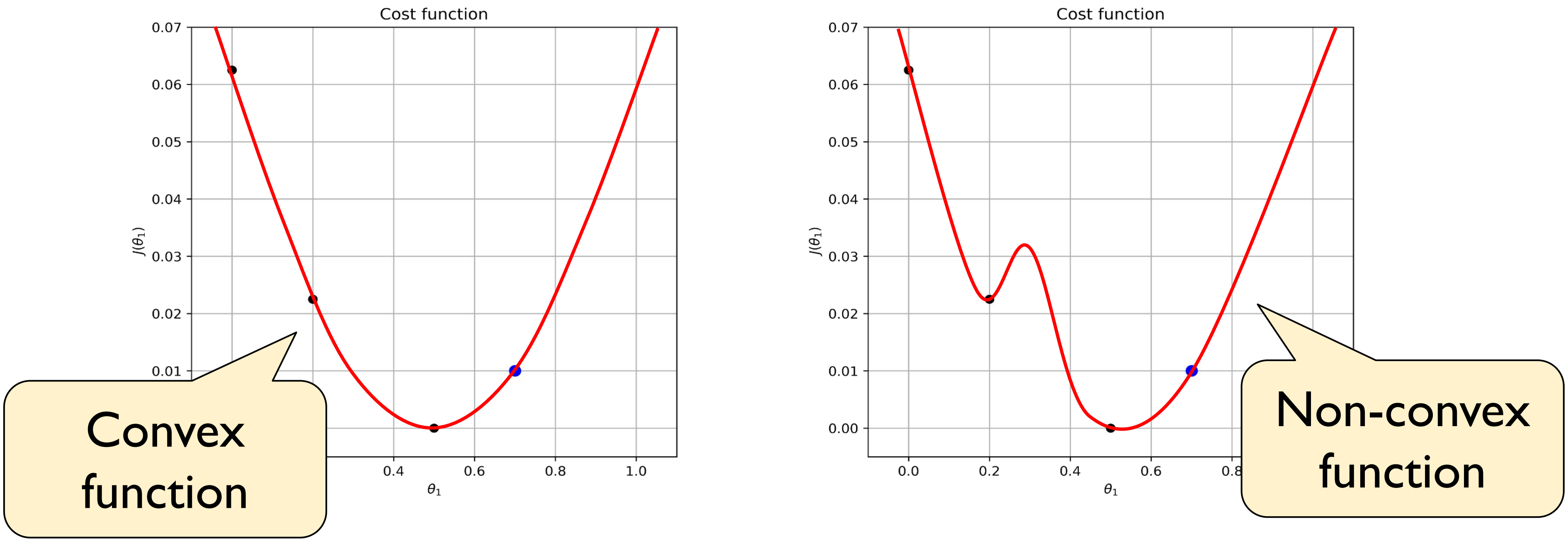
- Adjust the parameters θ t
  function $J(\theta)$
- Recap: Linear model

$$J(\theta) = \underbrace{\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2}_{\mathrm{MSE}(X, h_\theta)}$$



Training a logistic regression model

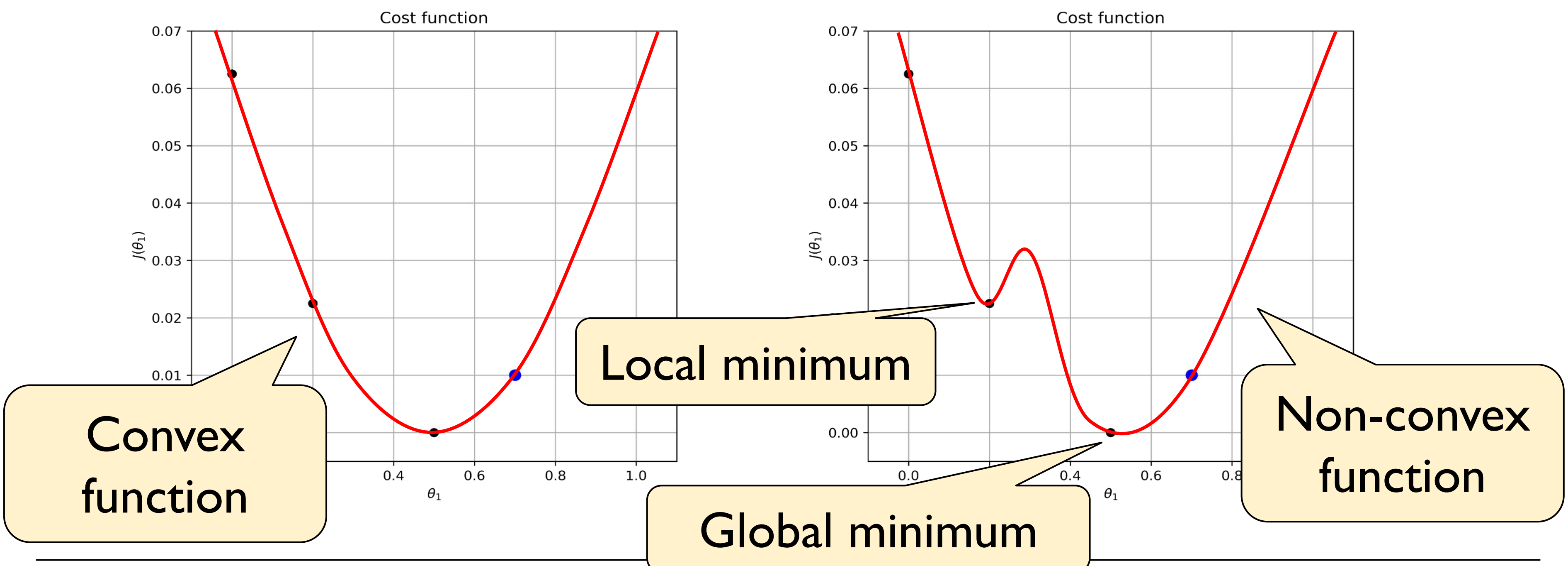
- Adjust the parameters θ to minimize the cost function $J(\theta)$
- <u>MSE isn't convex for logistic regression!</u>



Supervised Learning Algorithms
Logistic Regression

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- <u>MSE isn't convex for logistic regression!</u>



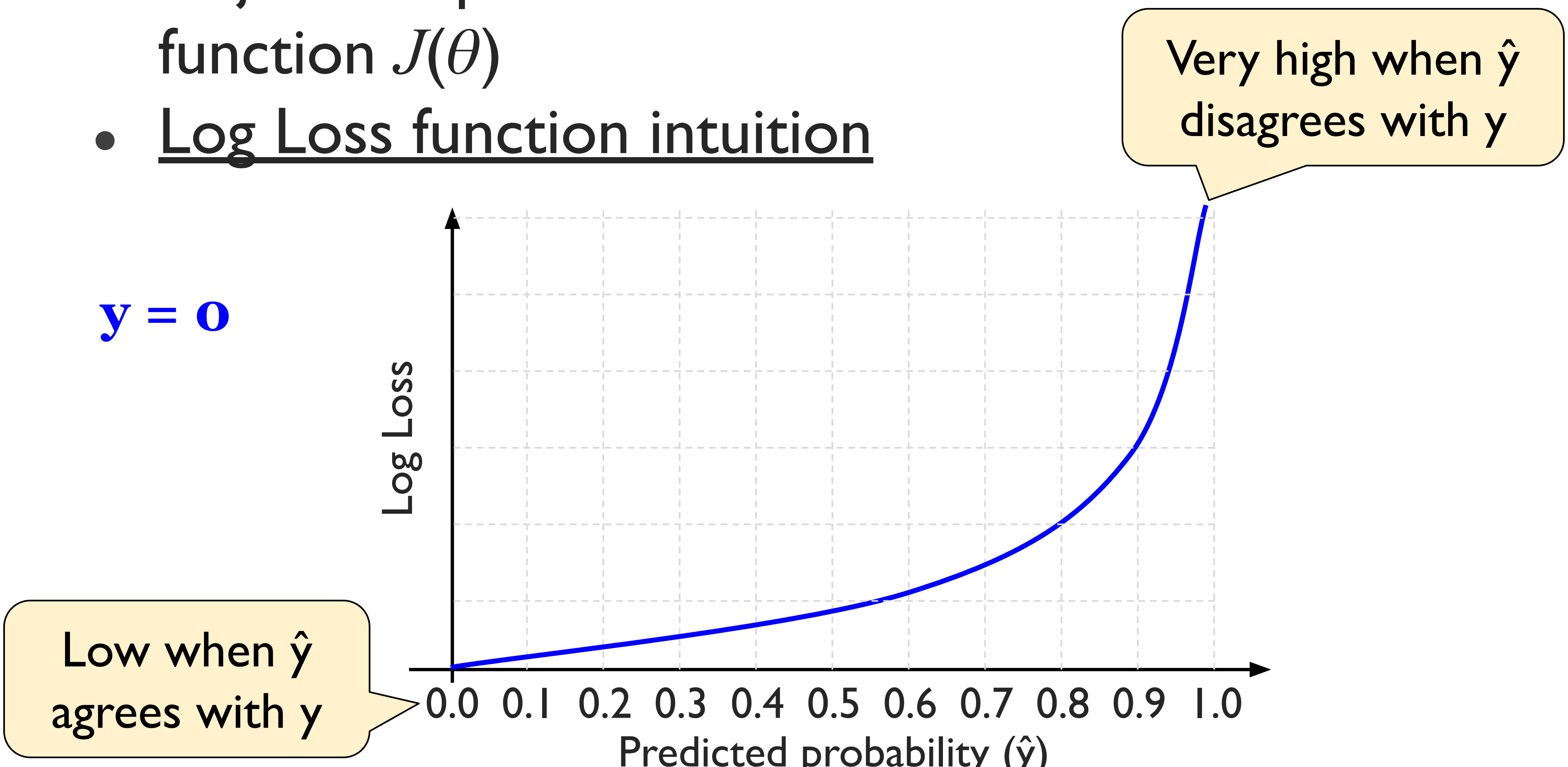
Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- <u>For logistic regression we use the log loss function!</u>

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- <u>Log Loss function intuition</u>

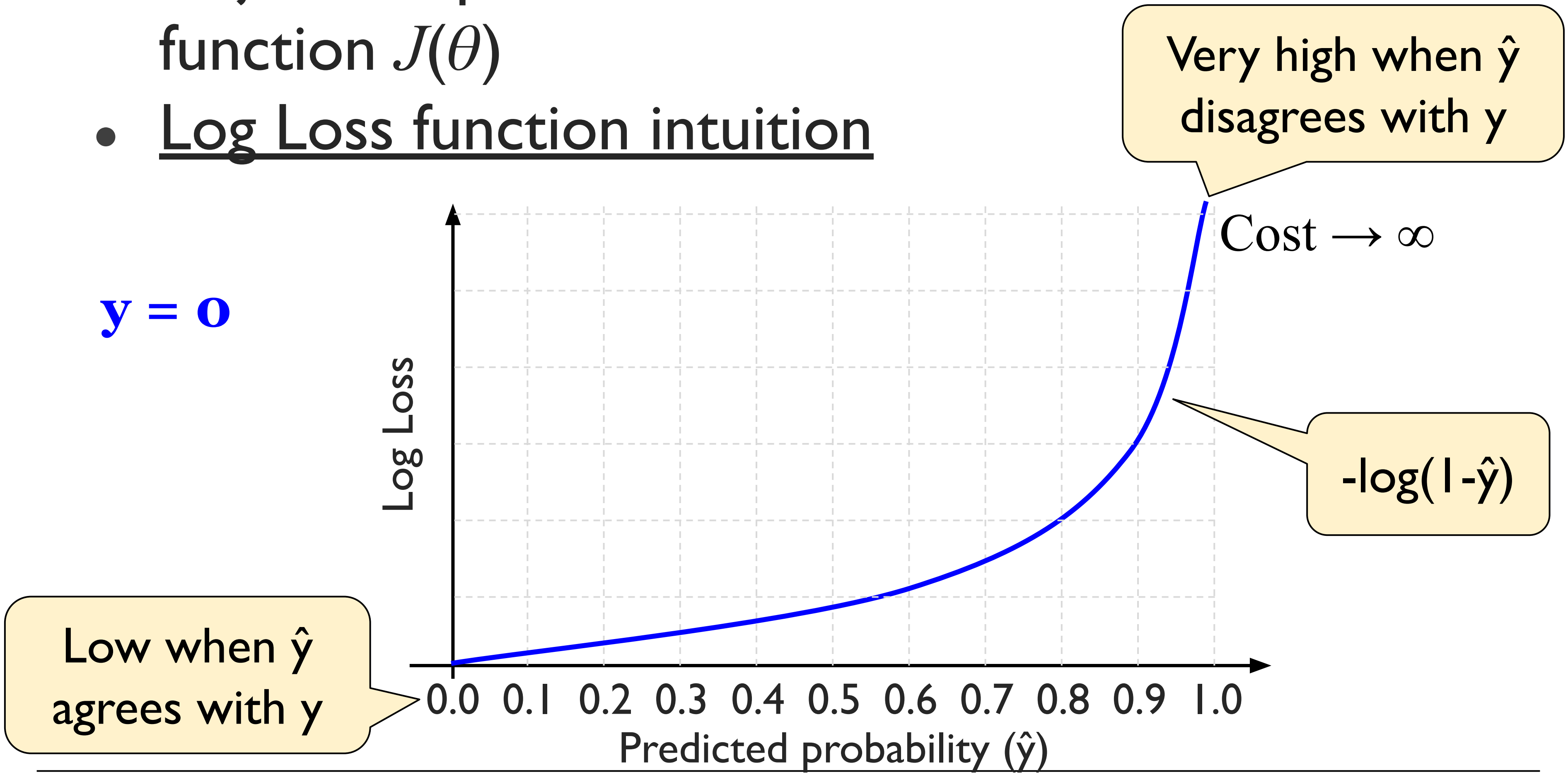
y = 0



Training a logistic regression model

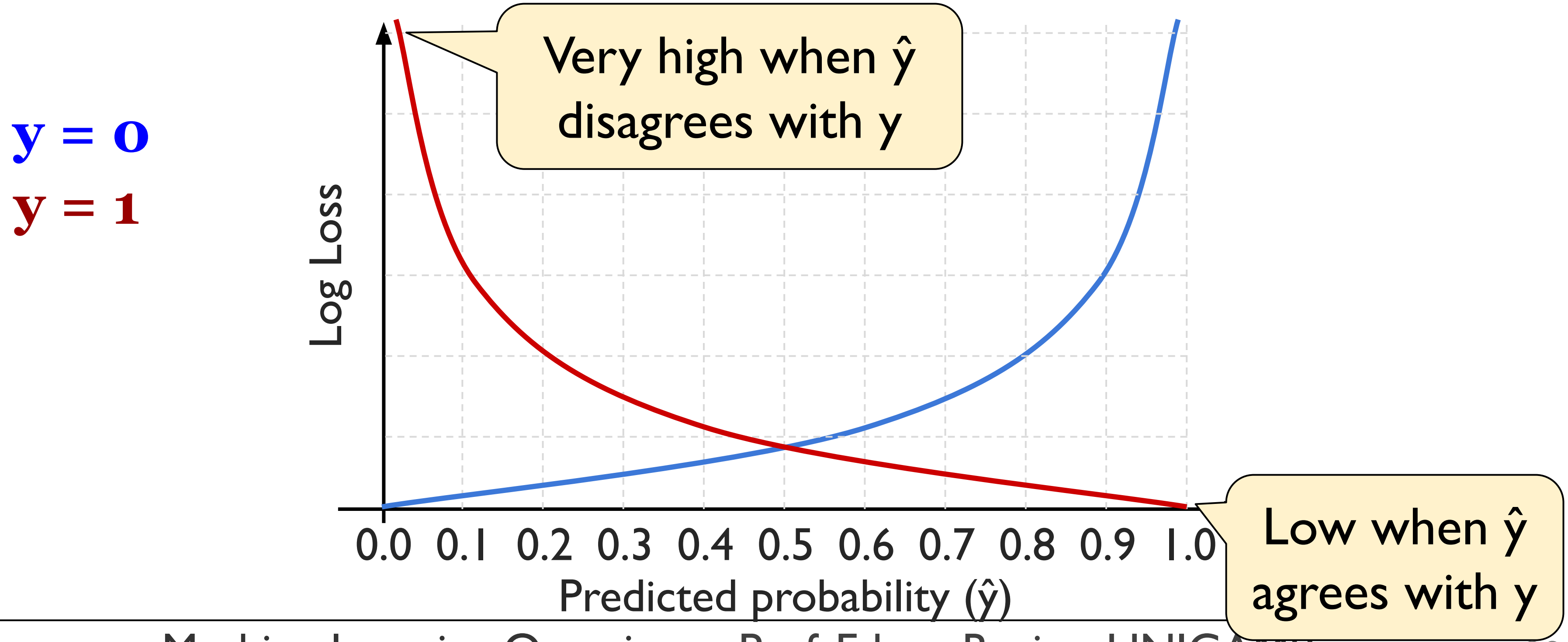
- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

y = 0



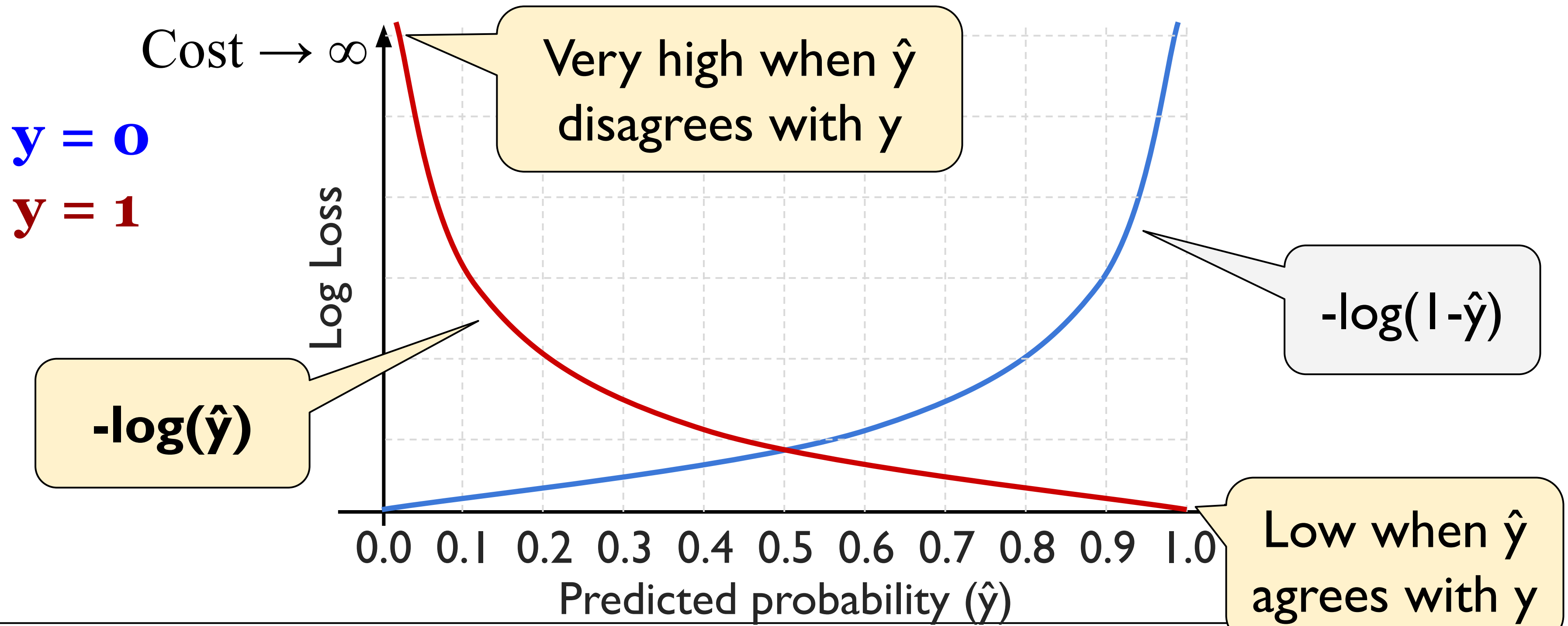
Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition



Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

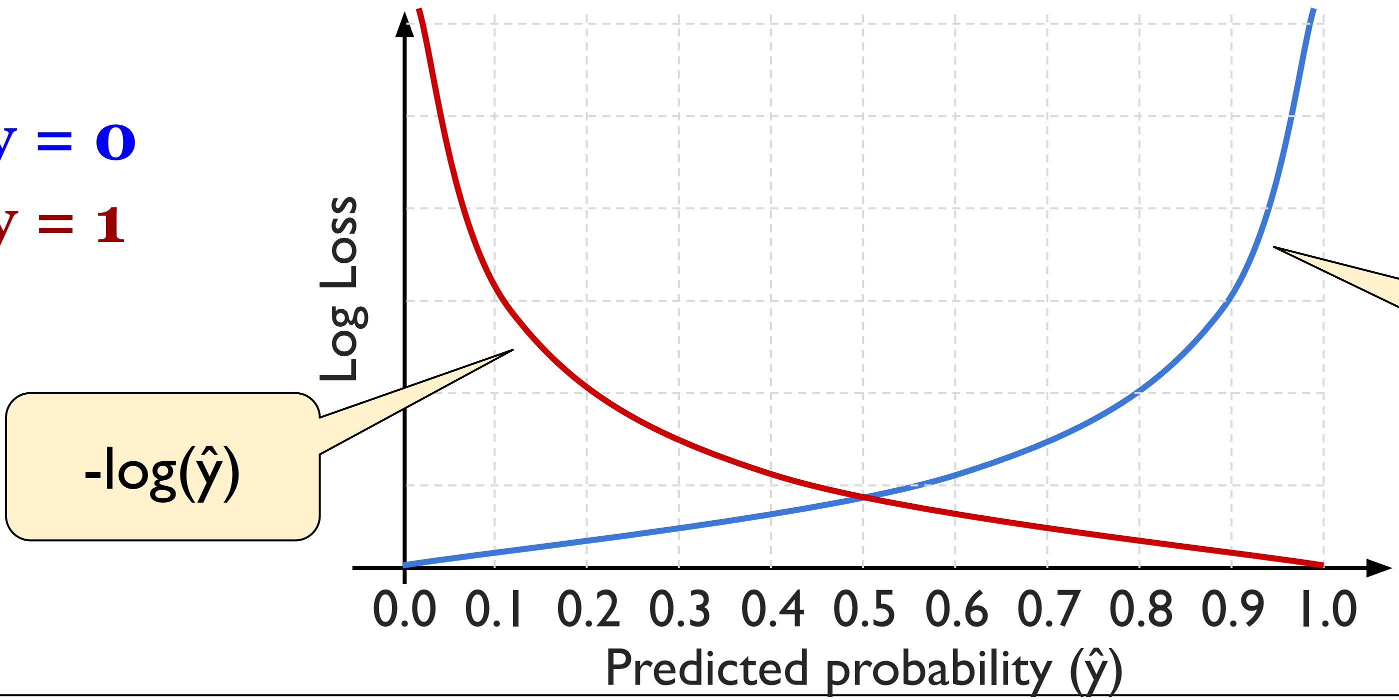


Training a logistic

- Adjust the param function $J(\theta)$
- Log Loss function intuiti

$$c(\theta) = \begin{cases} -\log(\, h_\theta(x)\,) & \text{if } y = 1 \\ -\log(\, 1 - h_\theta(x)\,) & \text{if } y = 0 \end{cases}$$

y = 0
y = 1



Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta) = \begin{cases} -\log(\ h_\theta(x)\ ) & \text{if } y = 1 \\ -\log(\ 1 - h_\theta(x)\ ) & \text{if } y = 0 \end{cases}$$

$$c(\theta) = -\log(\ h_\theta(x)\ ) \ .\ y - \log(\ 1\text{-}h_\theta(x)\ ) \ .\ (1\text{-}y)$$

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta) = \begin{cases} \underline{-\log(\, h_\theta(x)\,)} & \underline{\text{if } y = 1} \\ -\log(\, 1 - h_\theta(x)\,) & \text{if } y = 0 \end{cases}$$

$$c(\theta) = \boxed{-\log(\, h_\theta(x)} \,)\, .\, y - \log(\, 1 - h_\theta(x)\,)\, .\, (1 - y)$$

1

0

Supervised Learning Algorithms
Logistic Regression

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- <u>Log Loss function intuition</u>

$$c(\theta) = \begin{cases} -\log(\, h_\theta(x)\,) & \text{if } y = 1 \\ \mathbf{-\log(\, 1 - h_\theta(x)\,)} & \mathbf{if\ y = 0} \end{cases}$$

$$c(\theta) = \underbrace{\cancel{-\log(\, h_\theta(x)\,) \cdot y}}_{0} \; \boxed{\mathbf{-\log(\, 1 - h_\theta(x)\,)}} \cdot \underbrace{(1-y)}_{1}$$

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta) = -\log(h_\theta(x)) . y - \log(1-h_\theta(x)) . (1-y)$$

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta) = \text{-log}(\ h_\theta(x)\ )\ .\ \text{y} - \text{log}(\ 1\text{-}h_\theta(x)\ )\ .\ (1\text{-y})$$

$$J(\theta) = \underbrace{\frac{1}{\text{m}} \sum_{i=1}^{m} [\ \text{-log}(\ h_\theta(x^{(i)})\ )\ .\ y^{(i)} - \text{log}(\ 1\text{-}h_\theta(x^{(i)})\ )\ .\ (1\text{-}y^{(i)})\ ]}_{\text{Log Loss}}$$

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta, x, y) = -\log(\, h_\theta(x)\,)\,.\, y - \log(\, 1 - h_\theta(x)\,)\,.\,(1-y)$$

$$J(\theta) = \underbrace{\frac{1}{m}\sum_{i=1}^{m} c(\theta, x^{(i)}, y^{(i)})}_{\text{Log Loss}} + \underbrace{\boxed{\alpha\,.\,(\theta_1^{\,2} + \theta_2^{\,2} + \ldots + \theta_n^{\,2})}}_{l_2 \text{ (Ridge) regularization}}$$

Training a logistic regression model

- Adjust the parameters θ to minimize the cost function $J(\theta)$
- Log Loss function intuition

$$c(\theta, x, y) = -\log(\, h_\theta(x)\,)\,.\,y - \log(\, 1 - h_\theta(x)\,)\,.\,(1-y)$$

$$J(\theta) = \underbrace{\frac{1}{m} \sum_{i=1}^{m} c(\theta, x^{(i)}, y^{(i)})}_{\text{Log Loss}} + \underbrace{\boxed{\alpha\,.\,(|\theta_1| + |\theta_2| + \ldots + |\theta_n|)}}_{l_1 \text{ (Lasso) regularization}}$$

Multiclass classification

Examples:

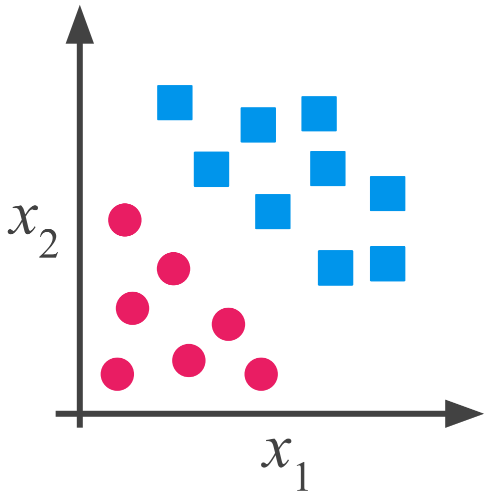
- Email tagging: Work, Friends, Family
  $y=1$ $y=2$ $y=3$
- HAR: Running, Walking, Jumping, …
  $y=1$ $y=2$ $y=3$

Multiclass classification

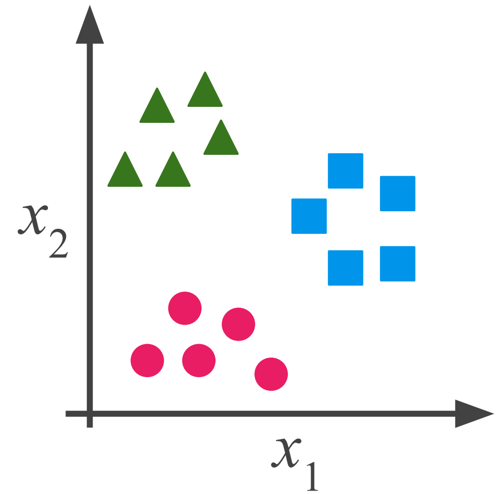
Examples:

- Email tagging: Work, Friends, Family
  $y = 1 \quad y = 2 \quad y = 3$
- HAR: Running, Walking, Jumping, …
  $y = 1 \quad y = 2 \quad y = 3$

Binary Classification

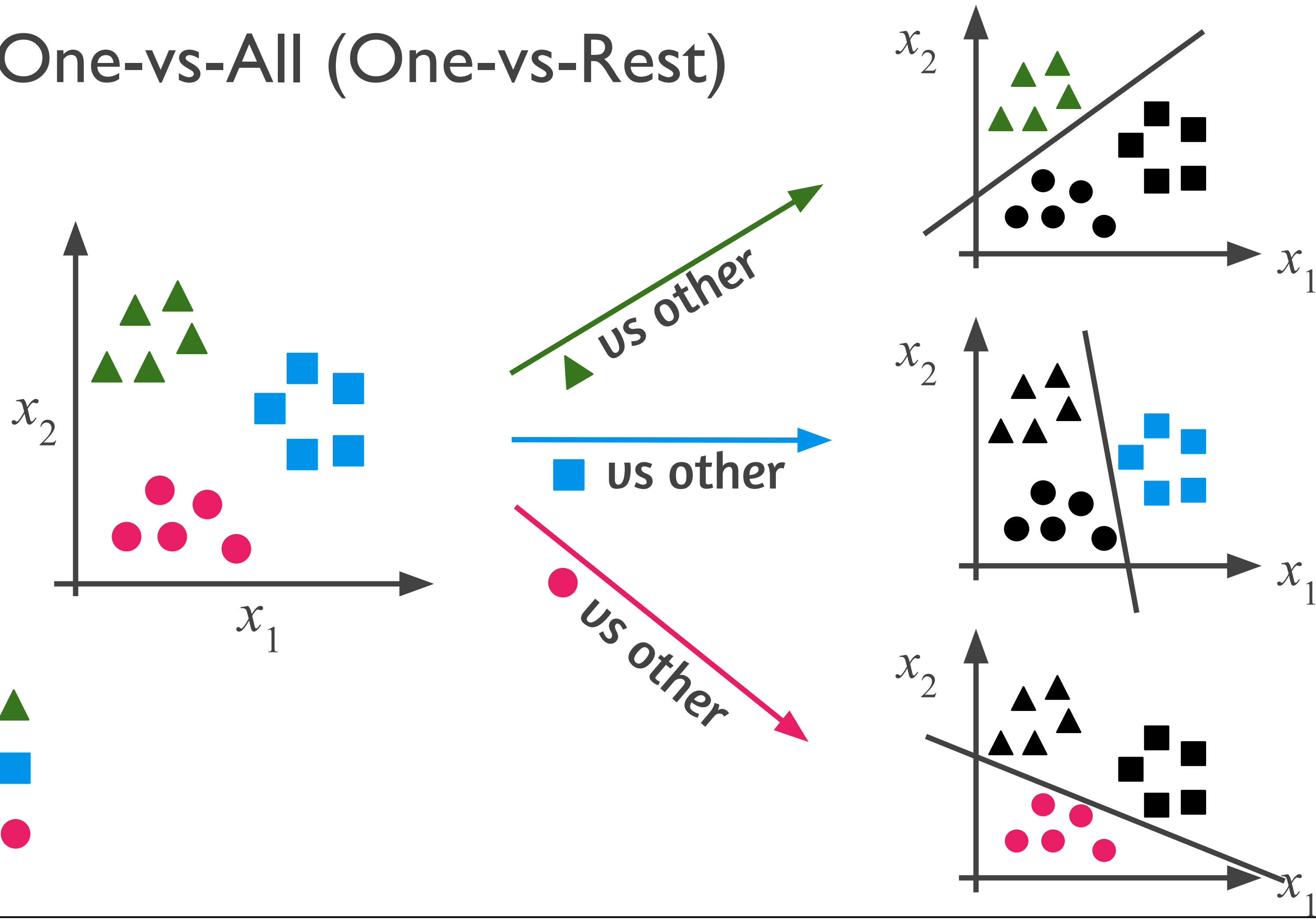


Multiclass Classification



Multiclass classification

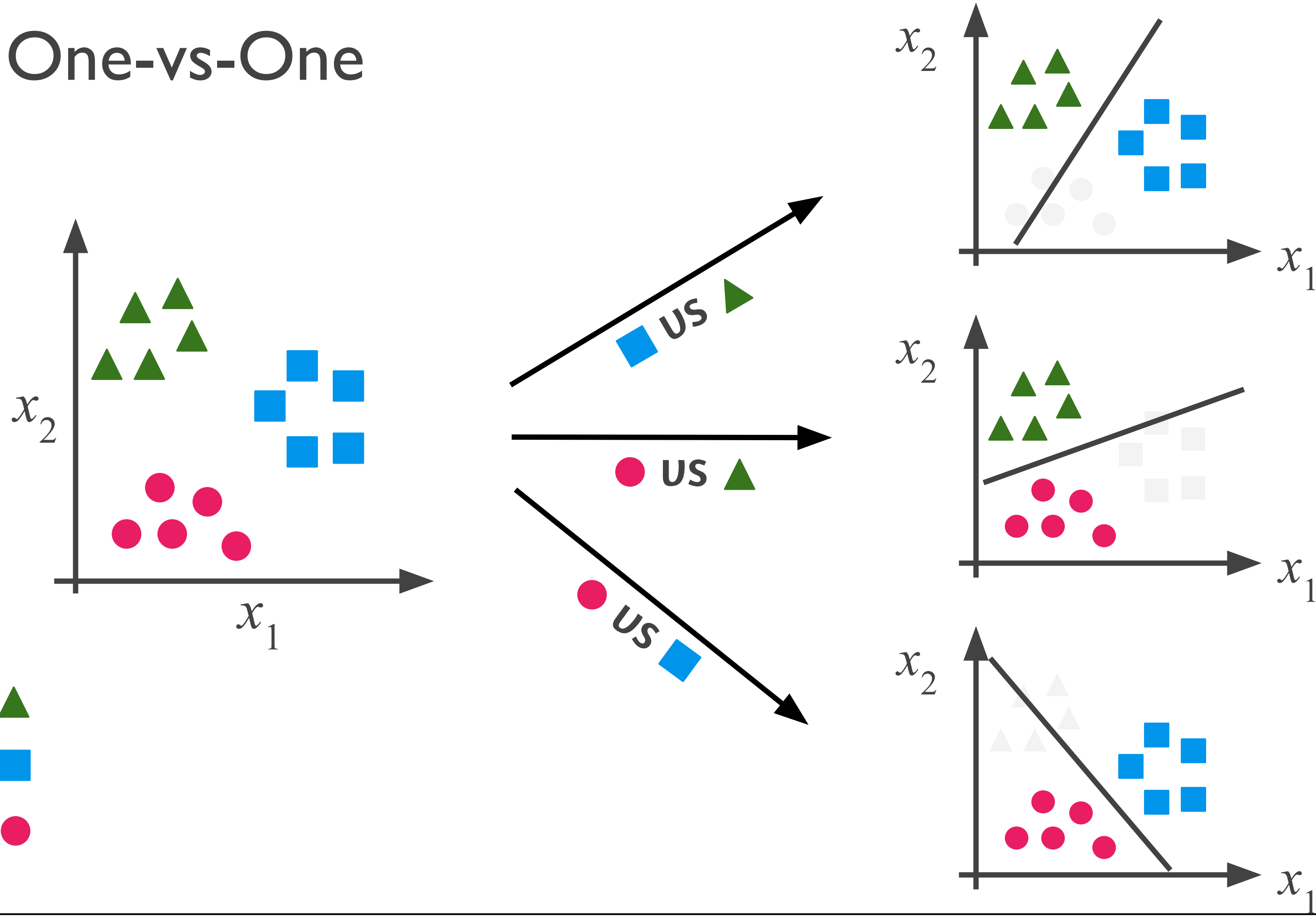
OvA: One-vs-All (One-vs-Rest)



Class 1: ▲
Class 2: ■
Class 3: ●

Multiclass classification

OvO: One-vs-One



Class 1: ▲
Class 2: ■
Class 3: ●

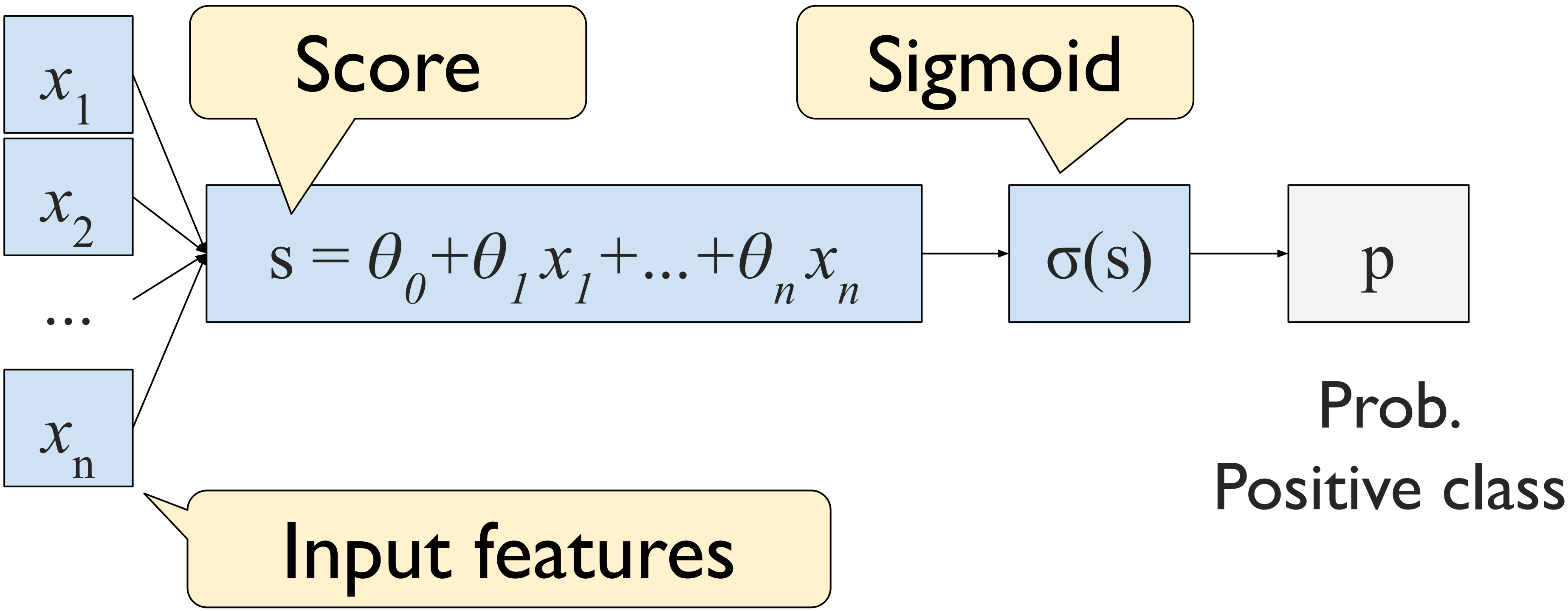
Softmax Regression

Supervised Learning Algorithms
Softmax Regression

Softmax Regression is a generalization of logistic regression that allows multiclass classification directly, without having to combine multiple binary classifiers

- It is also known as Multinomial Logistic Regression

Score on Logistic Regression



Softmax regression model

x_1

x_2

...

x_{n-1}

x_n

$$s^1 = \theta_0^{\,1} + \theta_1^{\,1} x_1 + \ldots + \theta_n^{\,1} x_n$$

s^1: Score for class 1

Softmax regression model

x_1

x_2

...

x_{n-1}

x_n

$$s^1 = \theta_0^{\,1} + \theta_1^{\,1} x_1 + \ldots + \theta_n^{\,1} x_n$$

$$s^2 = \theta_0^{\,2} + \theta_1^{\,2} x_1 + \ldots + \theta_n^{\,2} x_n$$

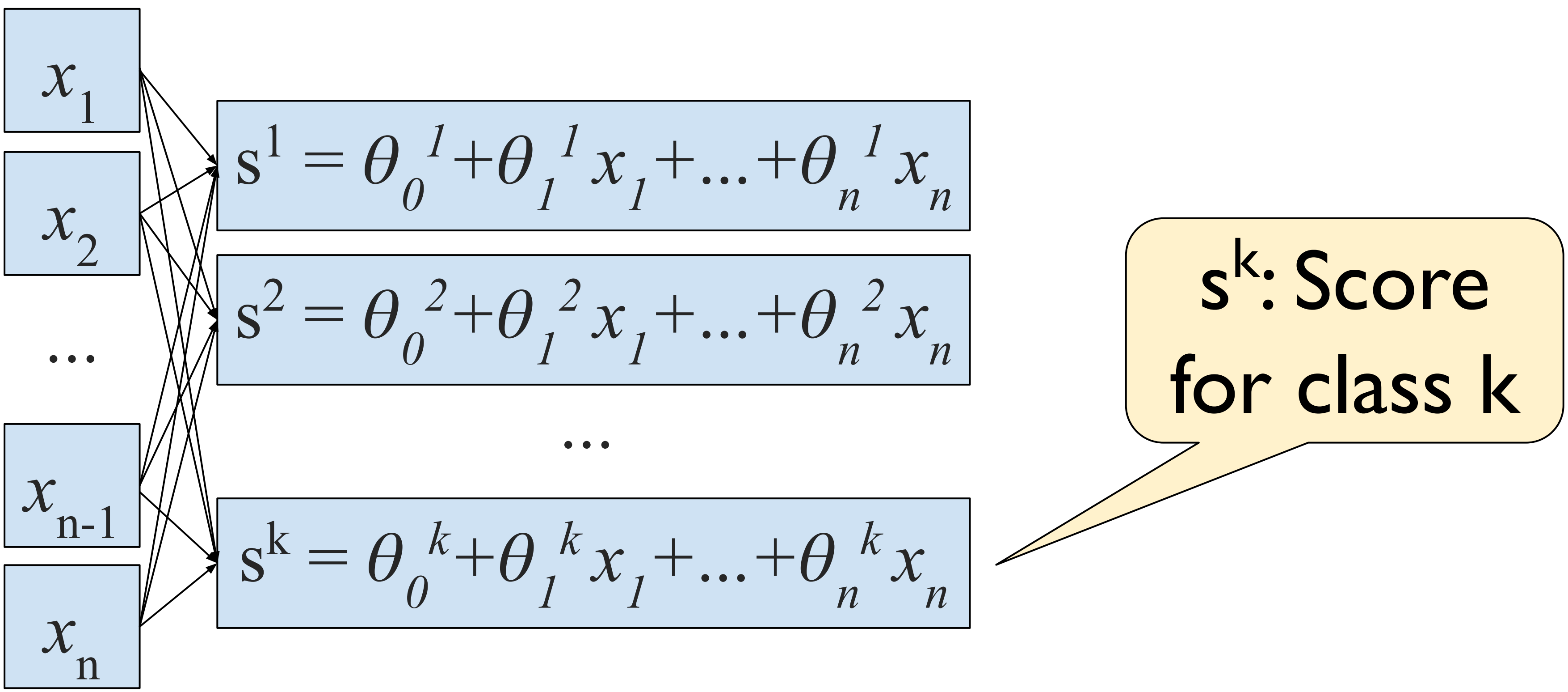
..

s^2: Score for class 2

Same input features, but different set of parameters (θ^2)

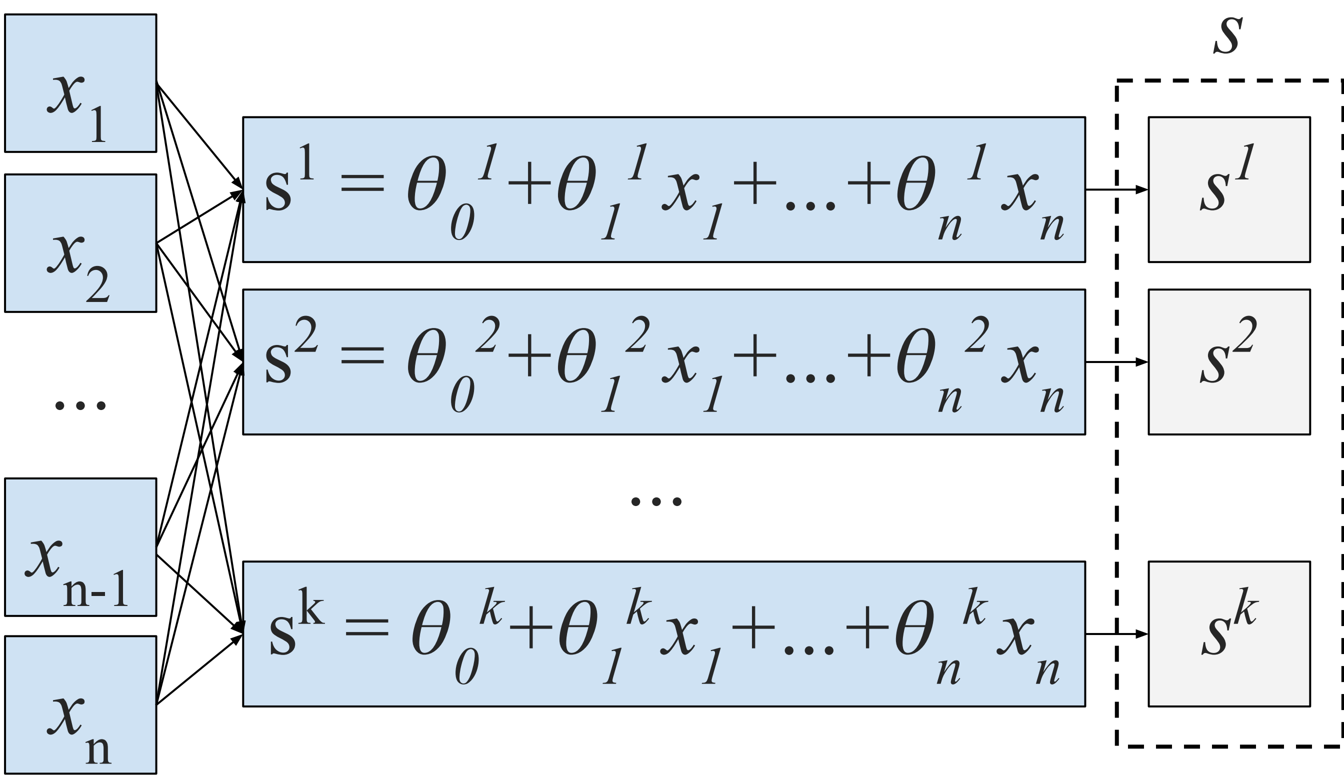
Supervised Learning Algorithms
Softmax Regression

Softmax regression model



Supervised Learning Algorithms
Softmax Regression

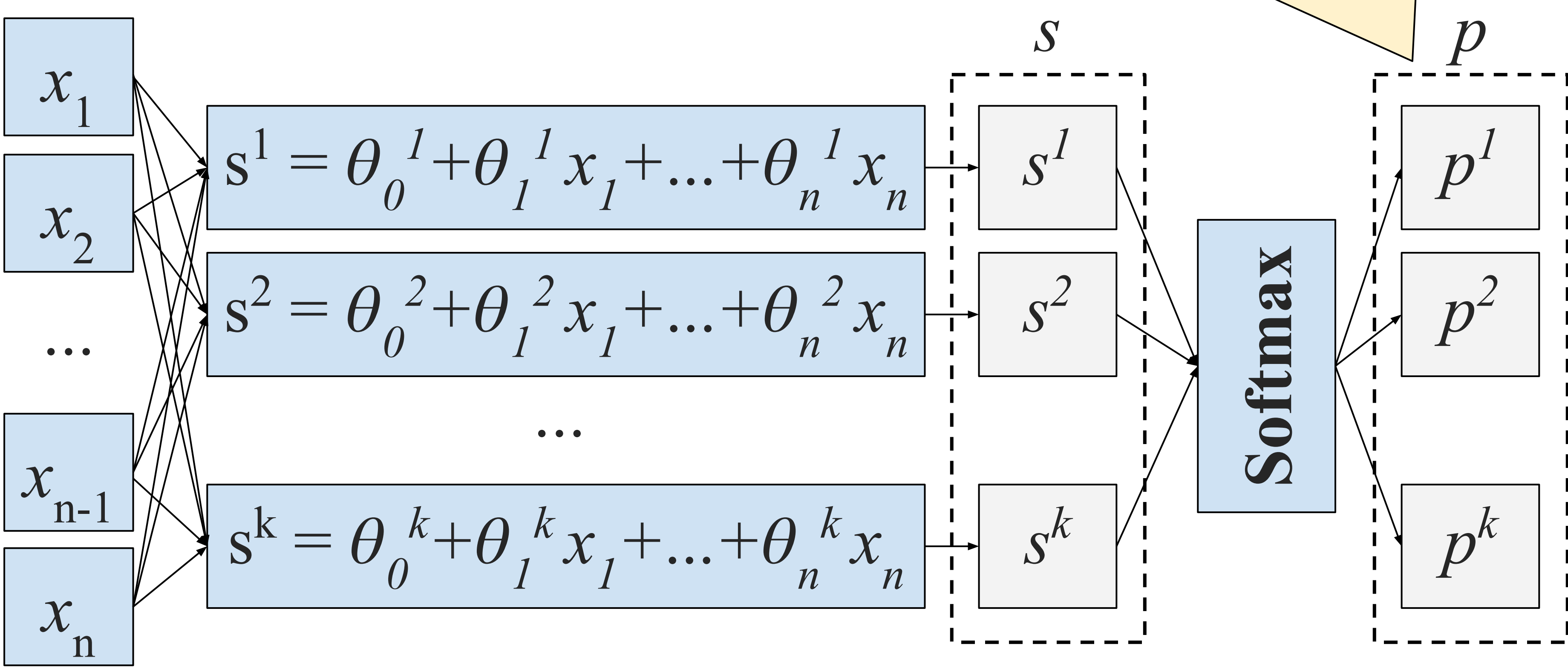
Softmax regression model



Supervised Learning Algorithms
Softmax Regression

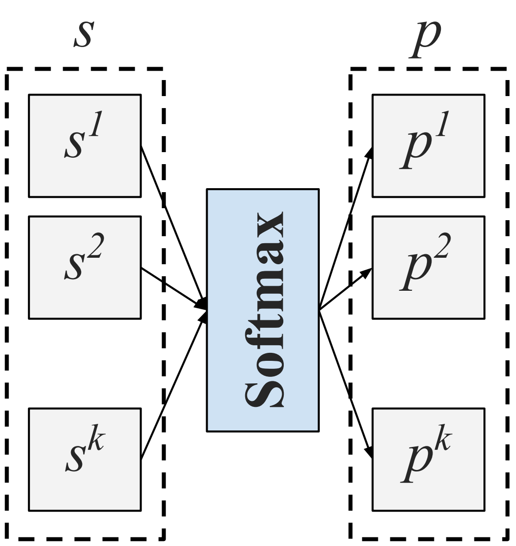
Softmax regression mode

p^i = Prob. x belongs to class i



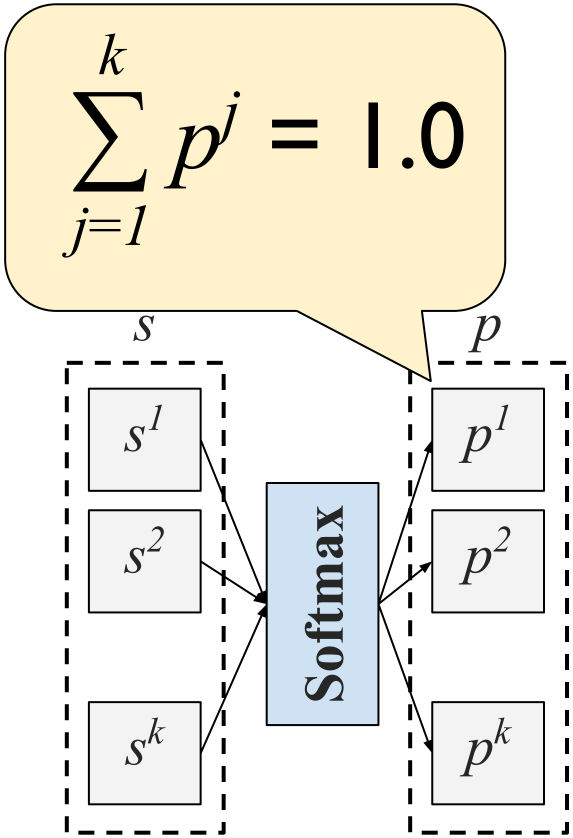
Softmax regression model

$$p^i = \text{Softmax}(s, i) = \frac{e^{s_i}}{\sum_{j=1}^{k} e^{s_j}}$$



Softmax regression model

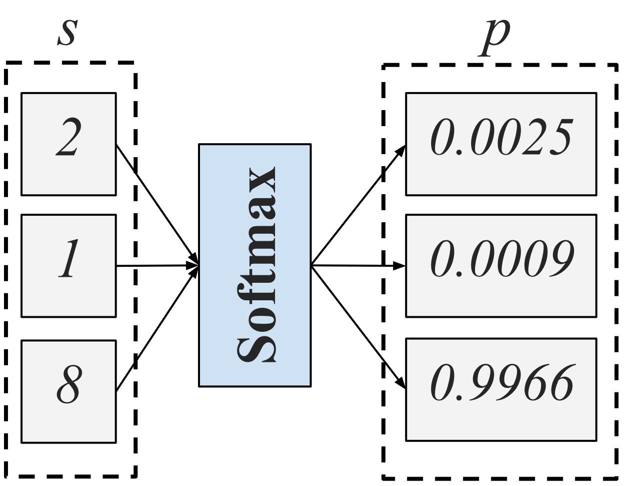
$$p^i = \mathrm{Softmax}(s, i) = \frac{e^{s_i}}{\sum_{j=1}^{k} e^{s_j}}$$



Softmax regression model

$$p^i = \mathrm{Softmax}(s, i) = \frac{e^{s_i}}{\sum_{j=1}^{k} e^{s_j}}$$

Example: k = 3



Softmax regression cost function

Cross entropy cost function:

$$J(\theta) = -\frac{1}{\mathrm{m}} \sum_{i=1}^{m} \sum_{k=1}^{K} [\, {y_k}^{(i)} \,.\, \log(h_\theta(x^{(i)})_k) \,]$$

Softmax regres

Cross entropy co

$y^{(i)}$ is a vector with k items

$y_k^{(i)}$: Prob. that i^{th} instance belongs to class k (either 1 or 0)

$$J(\theta) = - \frac{1}{\mathrm{m}} \sum_{i=1}^{m} \sum_{k=1}^{K} [\, y_k^{(i)} \,.\, \log(h_\theta(x^{(i)})_k) \,]$$

Softmax regres

$y^{(i)}$ is a vector with k items

$y_k^{(i)}$: Prob. that i^{th} instance belongs to class k (either 1 or 0)

Cross entropy co

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} [\, y_k^{(i)} \,.\, \log(h_\theta(x^{(i)})_k) \,]$$

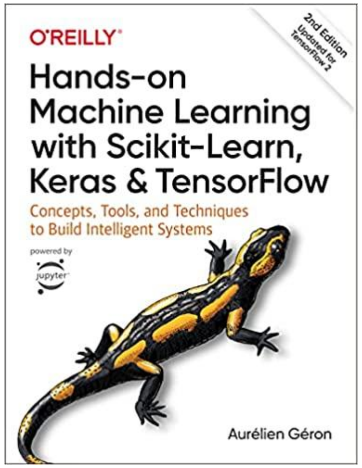
Example (K=3)

- $y^{(2)} \in$ class 2

$$\underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{y^{(2)}} \; . \; \underbrace{\begin{bmatrix} 0.9 \\ 0.02 \\ 0.08 \end{bmatrix}}_{h_\theta(x^{(2)})}$$

Supervised Learning Algorithms
Softmax Regression

References



- Aurelien Geron. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems - 2019
  - Chapter 4

- Logistic Regression on the scikit-learn website:
  - https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

Instituto de Computação
UNIVERSIDADE ESTADUAL DE CAMPINAS

Capacitação profissional em tecnologias de Inteligência Artificial

Machine Learning Overview

Prof. Edson Borin
https://www.ic.unicamp.br/~edson
Institute of Computing - UNICAMP

Video Idex

Video lessons	Length	Start	Length	Subtopics	
class-4.2: Logistic and Softmax regression	0:43:04	0:00:00	0:00:40	Logistic regression	Introduction
		0:00:40	0:01:32		Regression vs Classification recap
		0:02:12	0:02:21		Logistic Regression overview
		0:04:33	0:02:22		Logistic Regression model (sigmoid)
		0:06:55	0:07:54		Examples and decision boundary
		0:14:49	0:02:51		Training: Convex vs non-convex cost functions
		0:17:40	0:06:45		Training: Log loss cost function
		0:24:25	0:00:44		Regularization
		0:25:09	0:01:42		Multiclass classification
		0:26:51	0:02:35		One-vs-All
		0:29:26	0:02:26		One-vs-One
		0:31:52	0:00:30	Softmax regression	Definition
		0:32:22	0:00:52		Scores on Logistic Regression
		0:33:14	0:04:46		Softmax regression model function
		0:38:00	0:05:04		Softmax regression cost function (cross entropy)
		0:43:04			