Multiclass Classification: One vs All

Multiclass Classification

Logistic Regression

Multiclass classification

Email foldering/tagging: Work, Friends, Family, Hobby

Multiclass classification

Email foldering/tagging: Work, Friends, Family, Hobby

Multiclass classification

Email foldering/tagging: Work, Friends, Family, Hobby

Medical diagrams: Not ill, Cold, Flu

Multiclass classification

Email foldering/tagging: Work, Friends, Family, Hobby

$y=1$ $y=2$ $y=3$ $y=4$

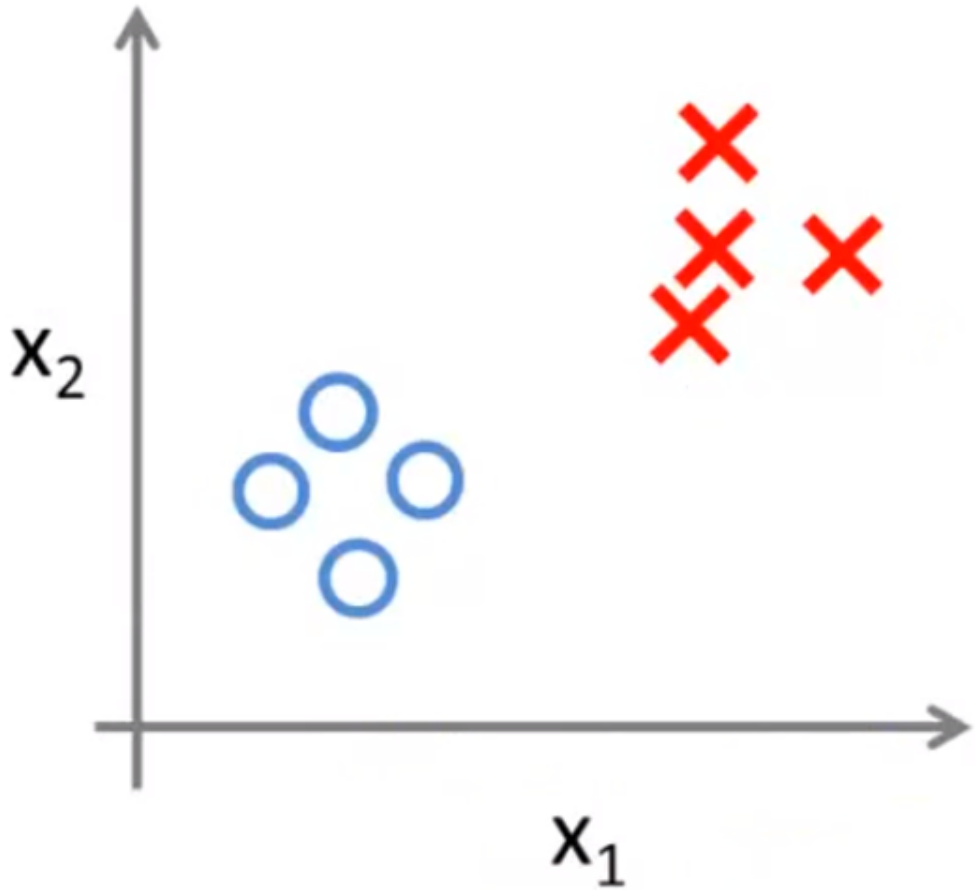
Medical diagrams: Not ill, Cold, Flu

$y=1$ 2 3

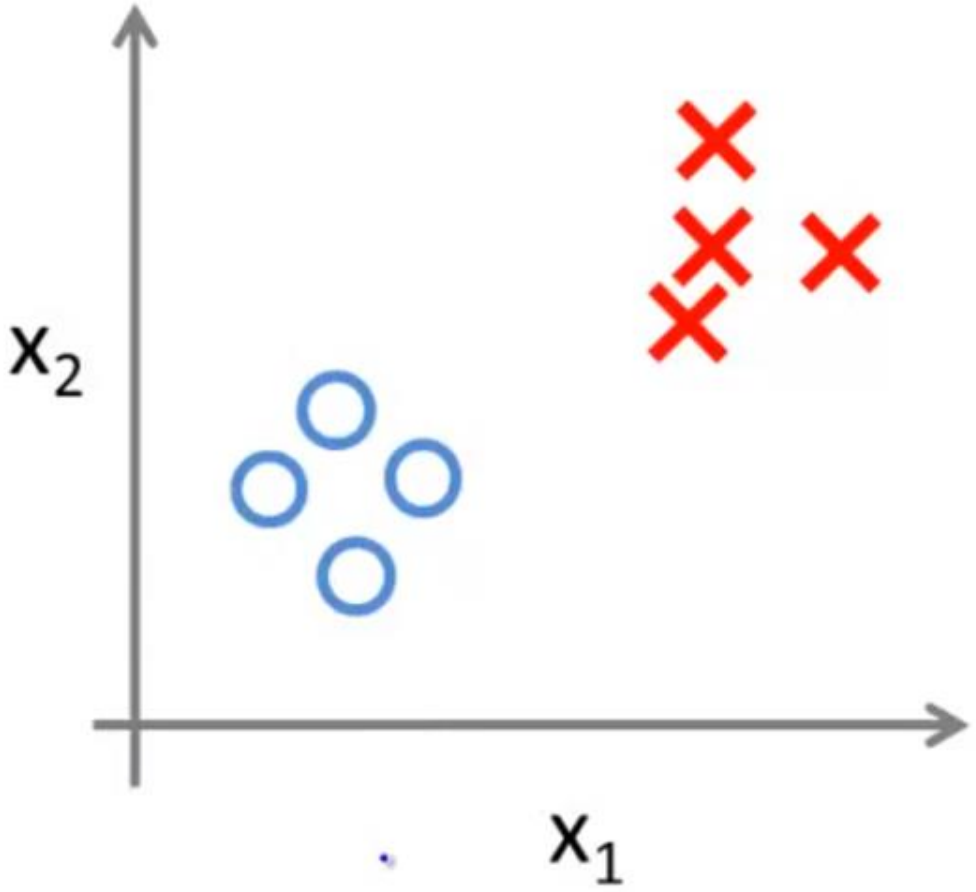
Weather: Sunny, Cloudy, Rain, Snow

$y=1$ 2 3 4

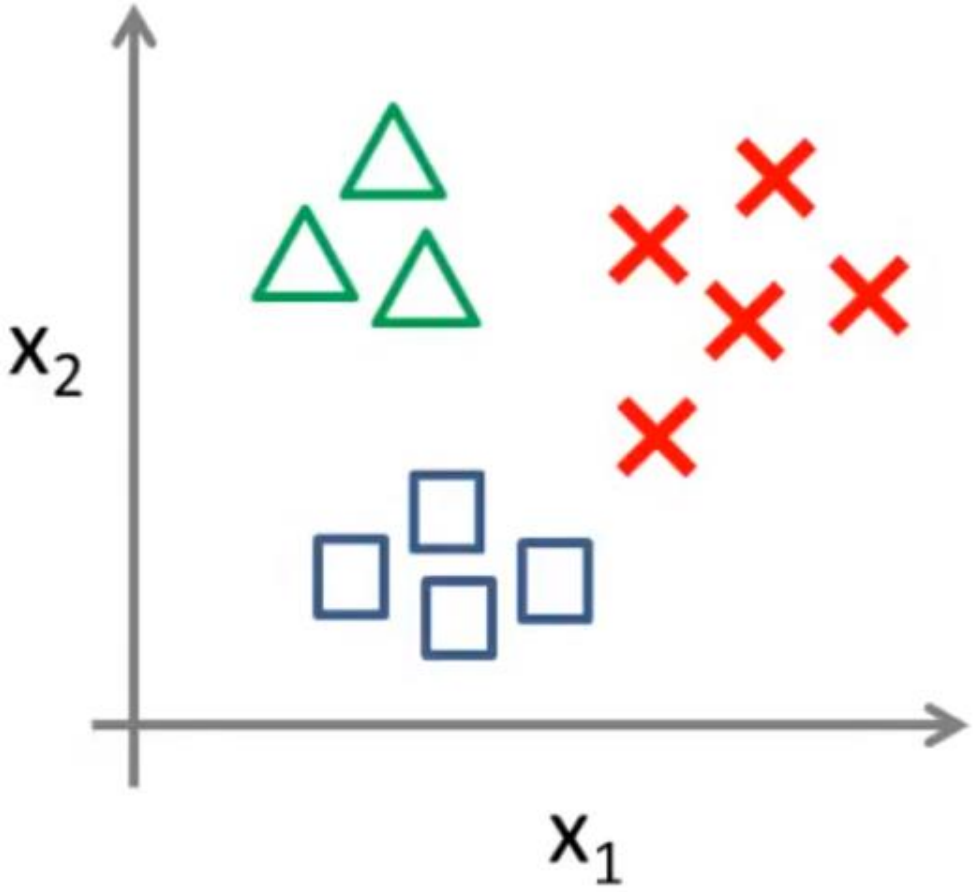
Binary classification:



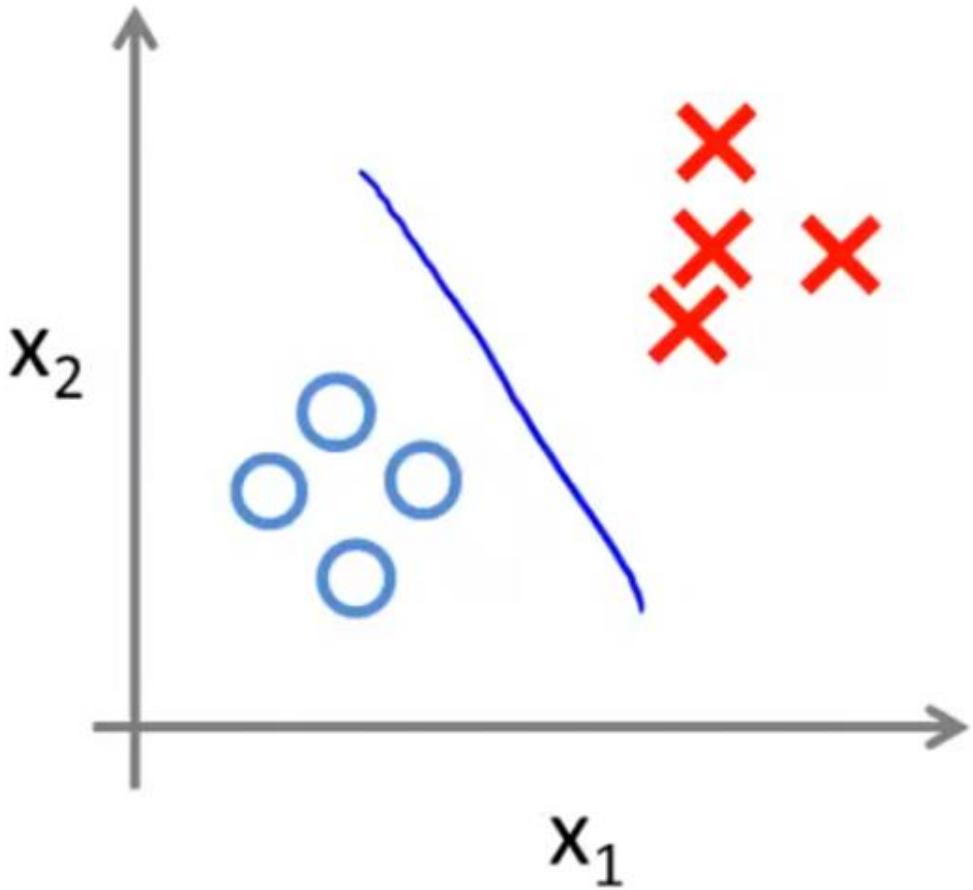
Binary classification:



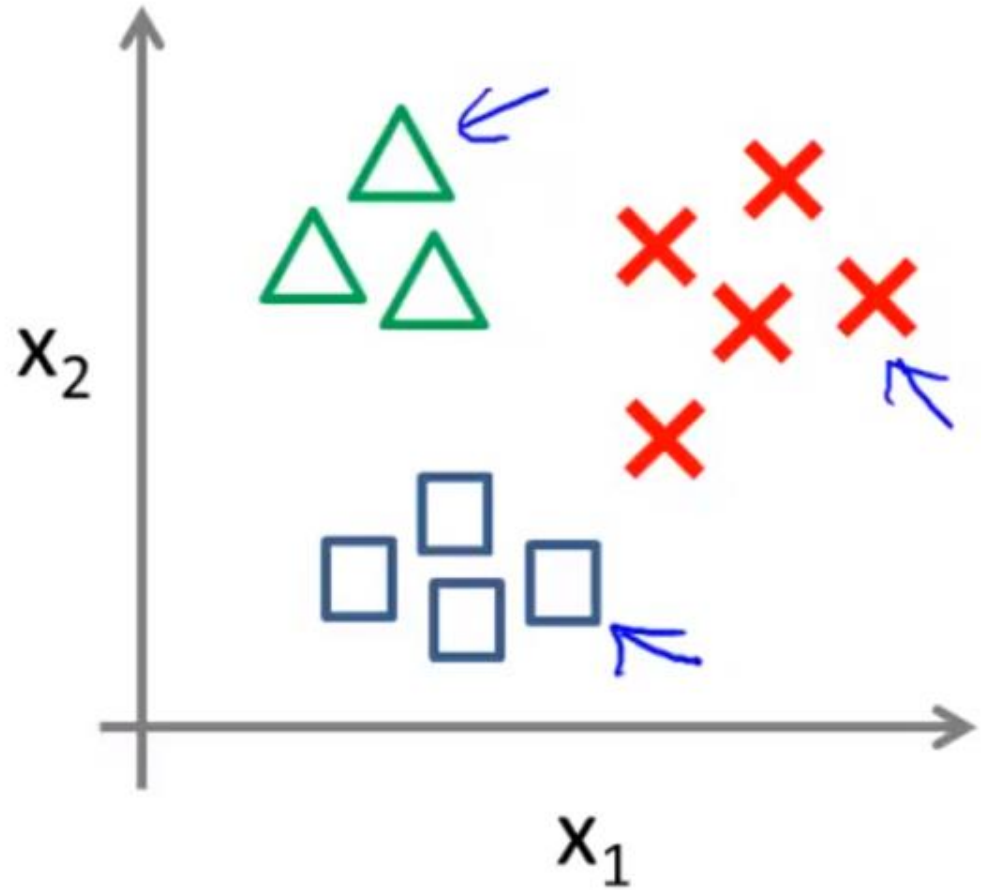
Multi-class classification:



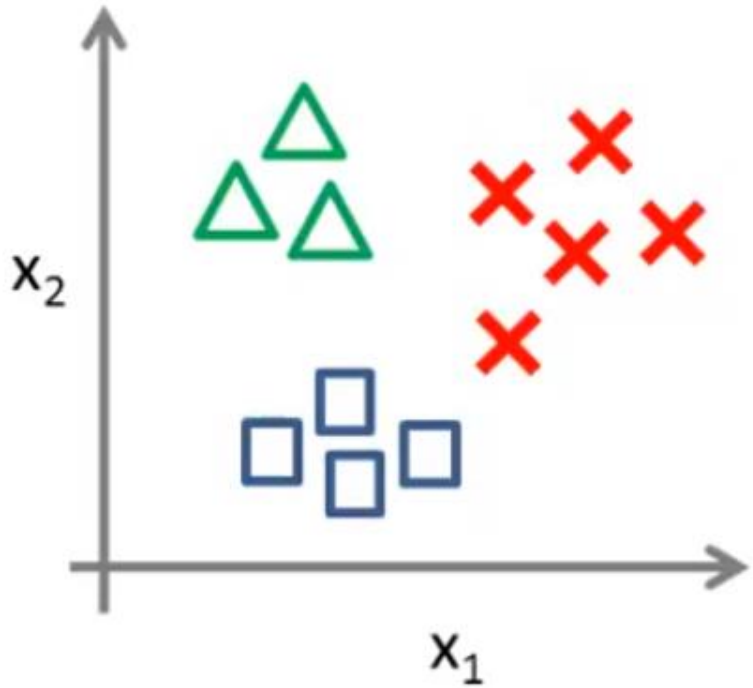
Binary classification:



Multi-class classification:



One-vs-all (one-vs-rest):

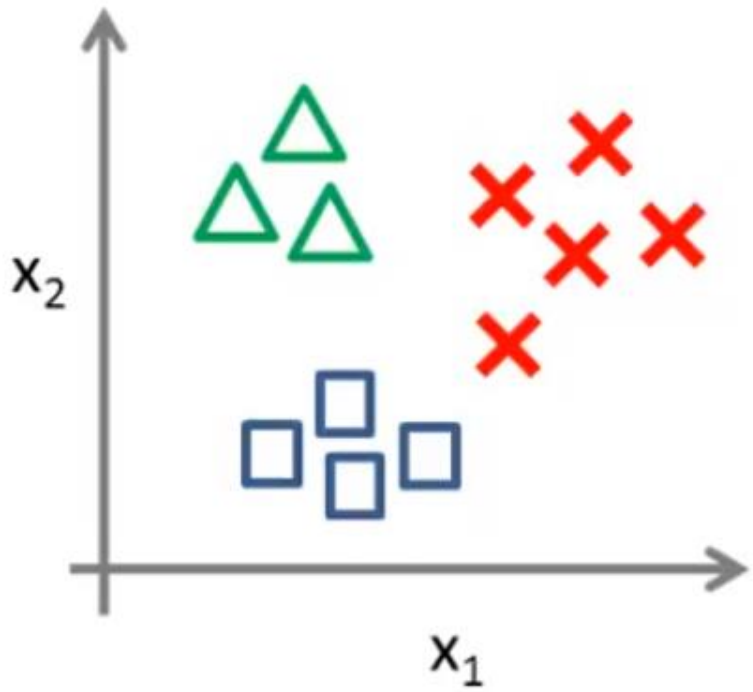


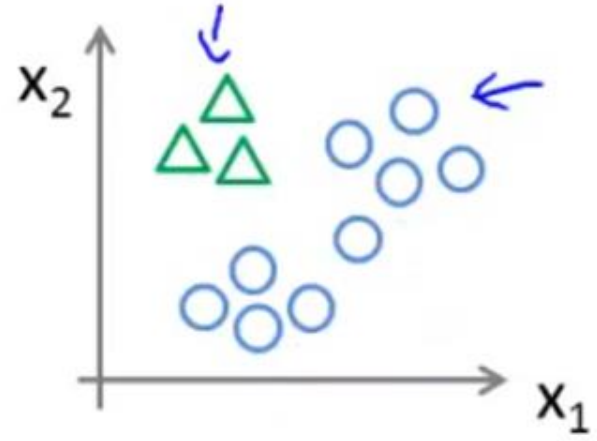
Class 1: △ ←

Class 2: □ ←

Class 3: × ←

One-vs-all (one-vs-rest):





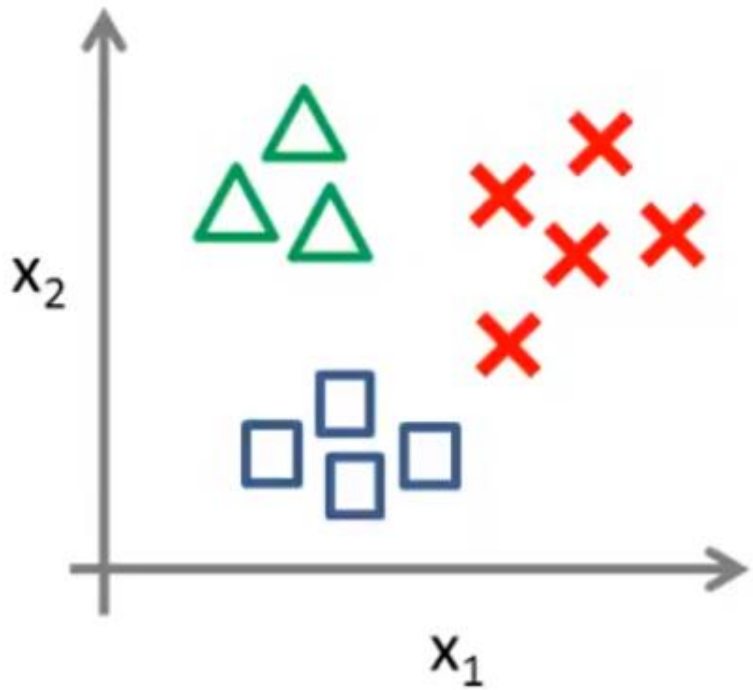
$h_\theta^{(1)}(x)$

Class 1: △ ←

Class 2: □ ←

Class 3: × ←

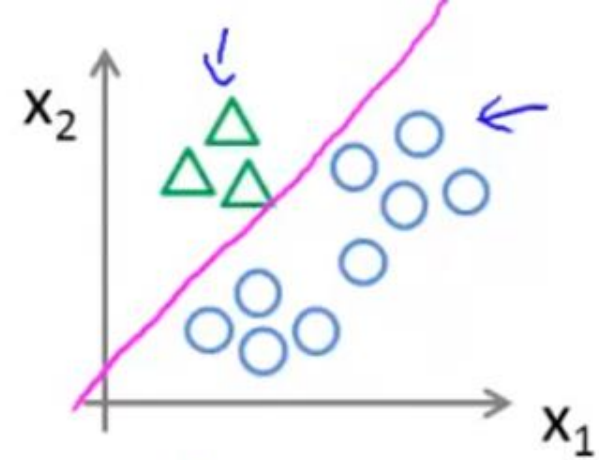
One-vs-all (one-vs-rest):

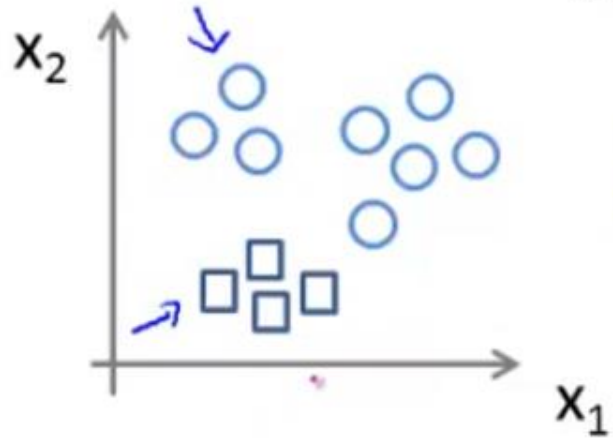


Class 1: △

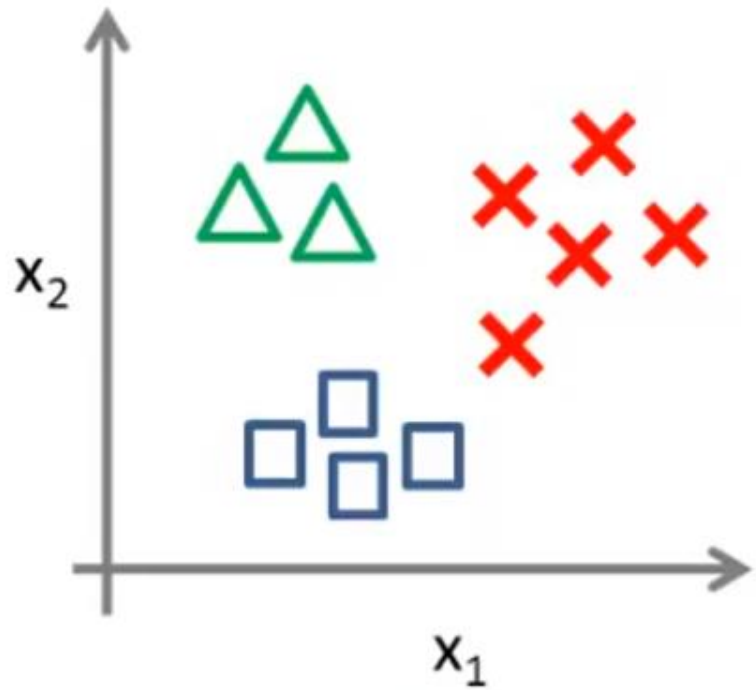
Class 2: □

Class 3: ×





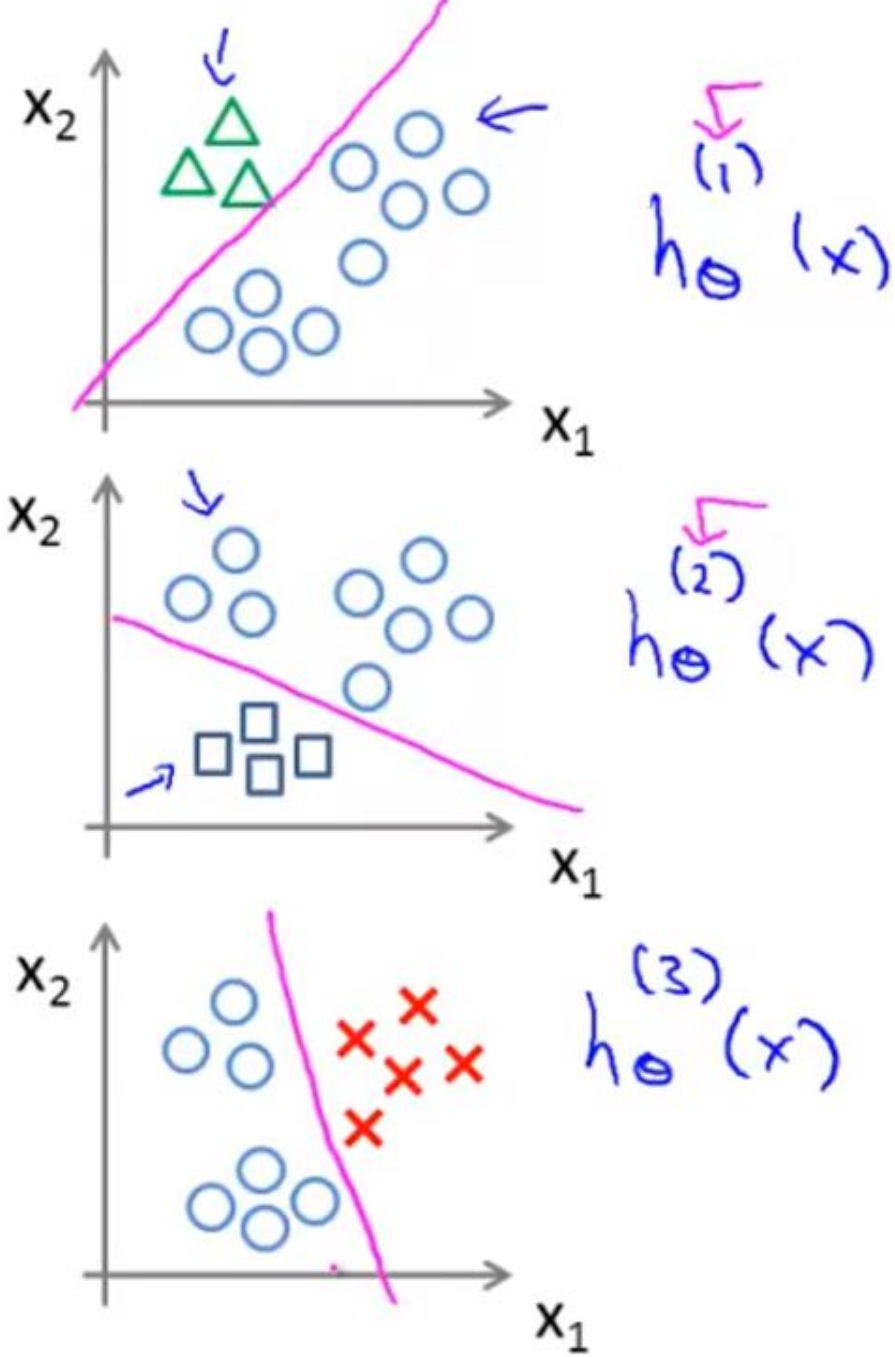
One-vs-all (one-vs-rest):



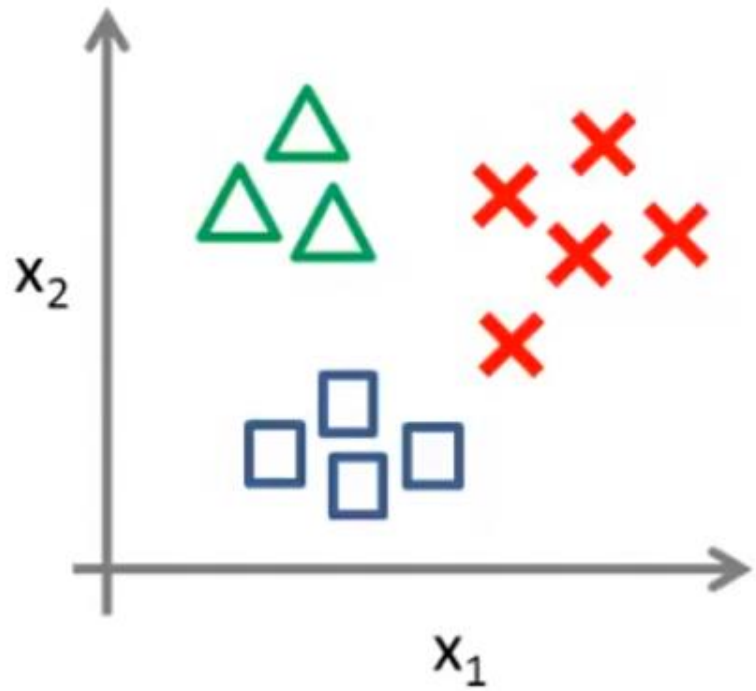
Class 1: △

Class 2: □

Class 3: ×



One-vs-all (one-vs-rest):

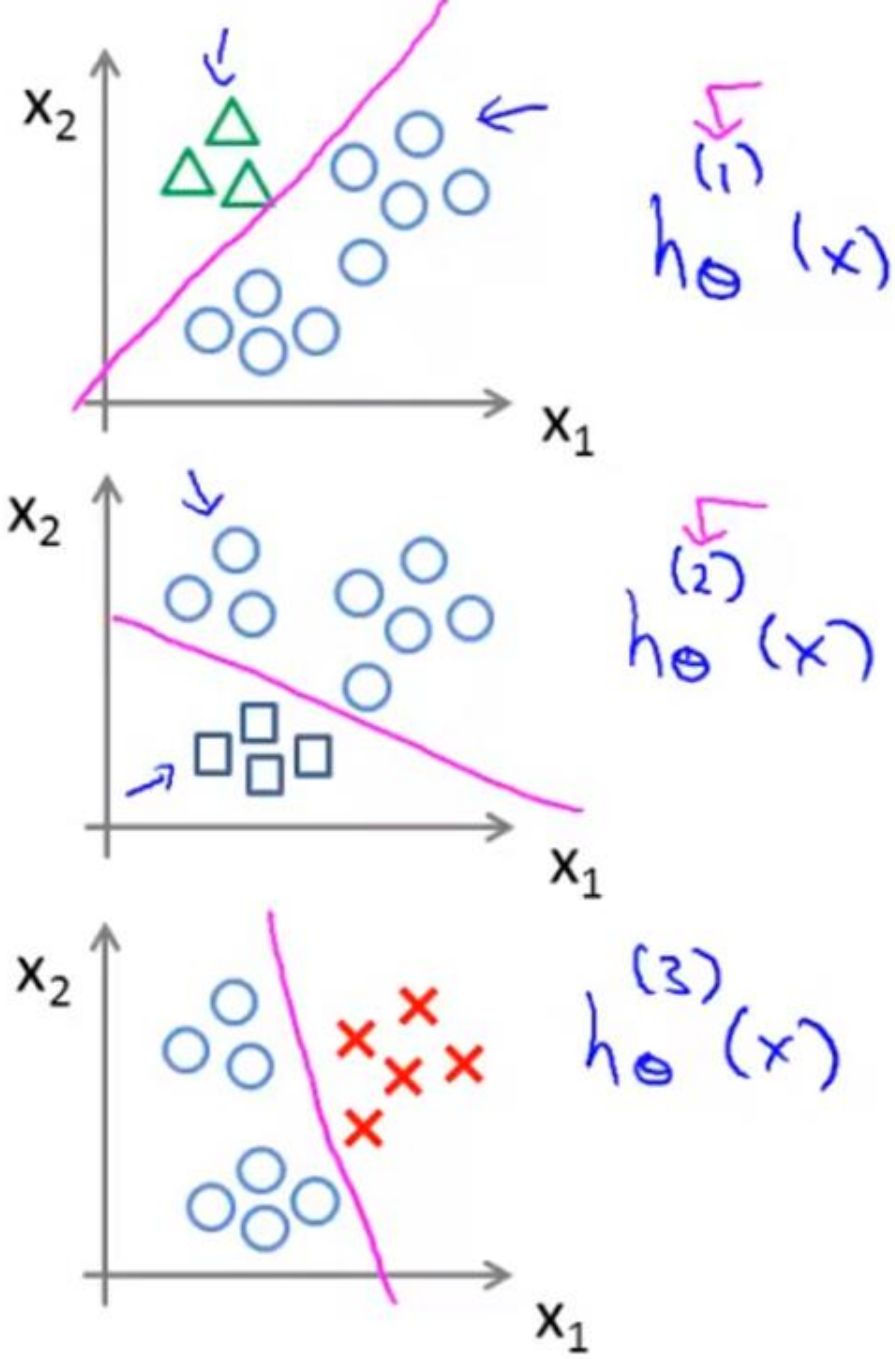


Class 1: △

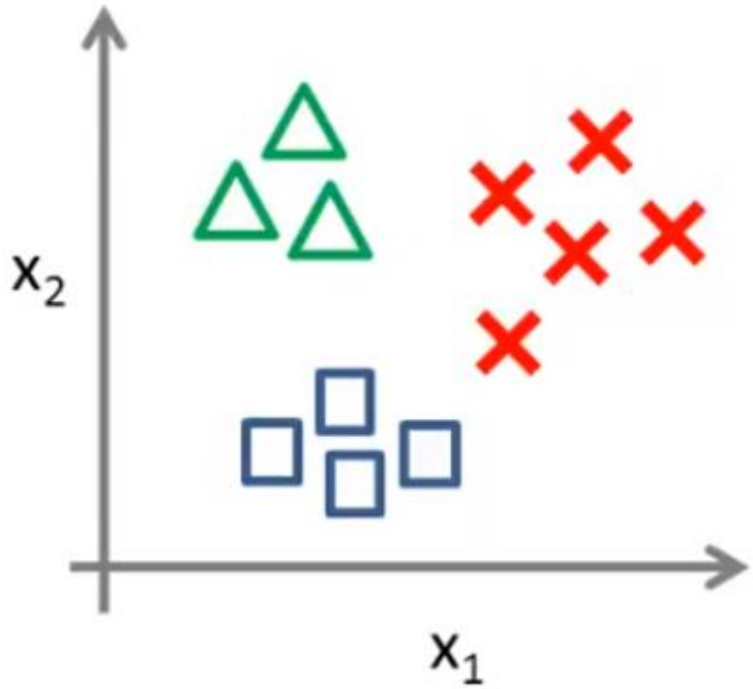
Class 2: □

Class 3: ×

$h_\theta^{(i)}(x) = P(y = i|x;\theta) \quad (i = 1, 2, 3)$



One-vs-all (one-vs-rest):

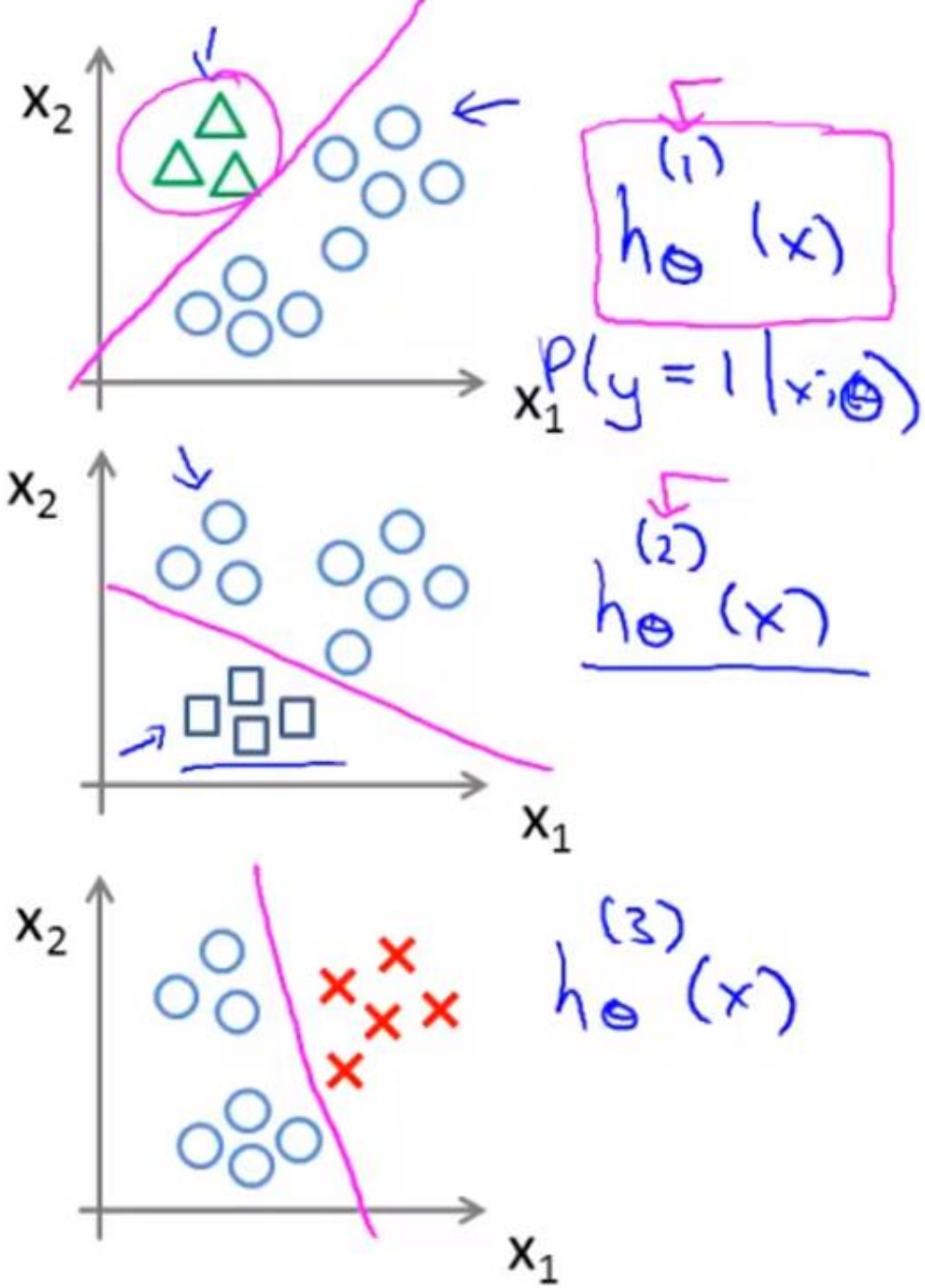


Class 1: △

Class 2: □

Class 3: ×

$$h_\theta^{(i)}(x) = P(y = i|x;\theta) \qquad (i = 1, 2, 3)$$



One-vs-all

Train a logistic regression classifier $h_\theta^{(i)}(x)$ for each class i to predict the probability that $y = i$.

On a new input x, to make a prediction, pick the class i that maximizes

$$\max_i h_\theta^{(i)}(x)$$

Exercise

- Suppose you have a multi-class classification problem with *k* classes ($y \in \{1,2,\ldots,k\}$). Using the 1-vs.-all method, how many different logistic regression classifiers will you end up training?
  - K-1
  - K
  - K+1
  - Approximately $\log_2(k)$