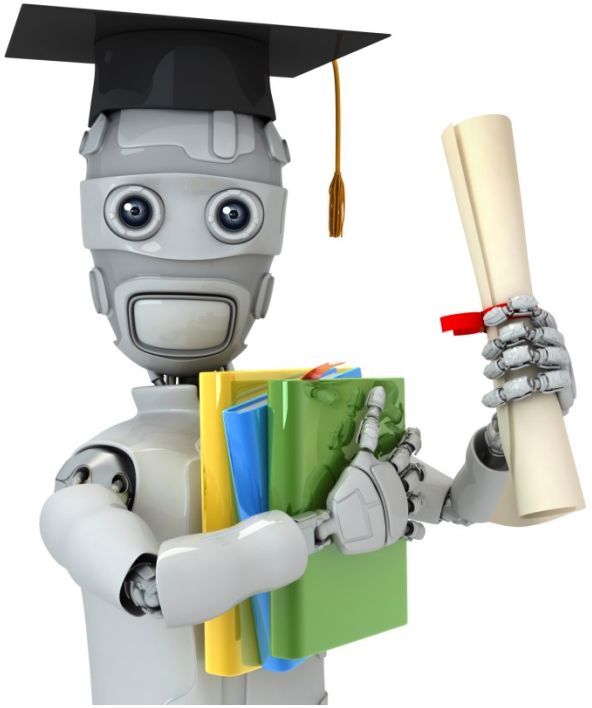


Machine Learning

Machine learning system design

Error metrics for skewed classes

Cancer classification example

Train logistic regression model $h_\theta(x)$. ($y = 1$ if cancer, $y = 0$ otherwise)

Lr

Find that you got 1% error on test set.
(99% correct diagnoses)

neste caso o 1% erro ja nao parece impresionante

1% ~= 0.5%

Only 0.50% of patients have cancer.

→ skewed classes. temos mais exemplos de uma classe do que outra

```
function y = predictCancer(x)
  → y = 0; %ignore x!
return
```

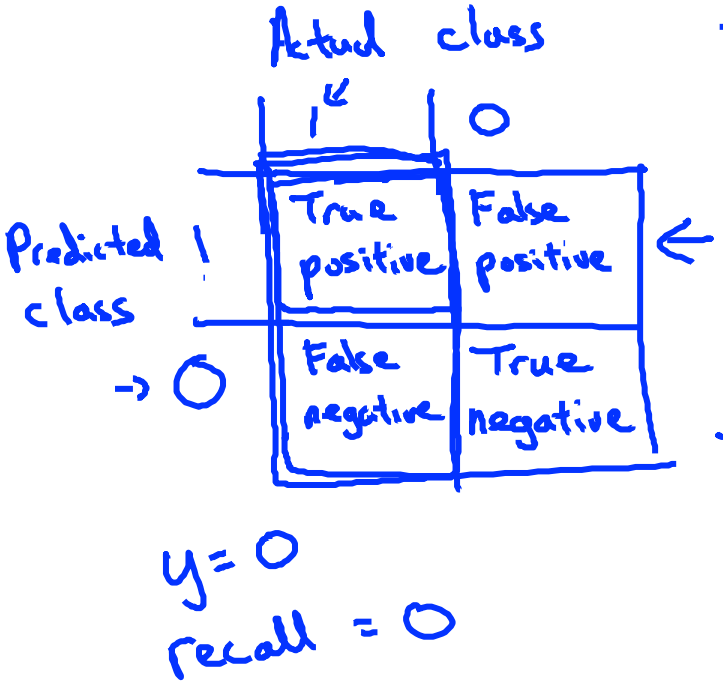
0.5% error

→ 99.2% accuracy (0.8% error)

→ 99.5% accuracy (0.5% error)

Precision/Recall

$y = 1$ in presence of rare class that we want to detect



→ **Precision**

(Of all patients where we predicted $y = 1$, what fraction actually has cancer?)

$$\frac{\text{True positives}}{\#\text{predicted positive}} = \frac{\text{True positive}}{\text{True pos + False pos}}$$

→ **Recall**

(Of all patients that actually have cancer, what fraction did we correctly detect as having cancer?)

$$\frac{\text{True positives}}{\#\text{actual positives}} = \frac{\text{True positives}}{\text{True pos + False neg}}$$