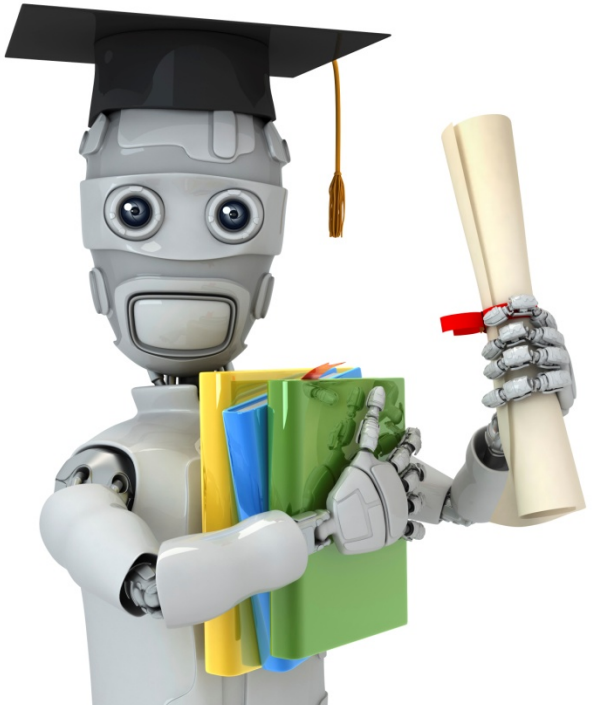


Machine Learning

Machine learning system design

Trading off precision and recall

Trading off precision and recall

- Logistic regression: $0 \leq h_\theta(x) \leq 1$

  Predict 1 if $h_\theta(x) \geq 0.5$ ~~0.7~~ ~~0.9~~ 0.3

  Predict 0 if $h_\theta(x) < 0.5$ ~~0.7~~ ~~0.9~~ 0.3

- Suppose we want to predict $y = 1$ (cancer) only if very confident.

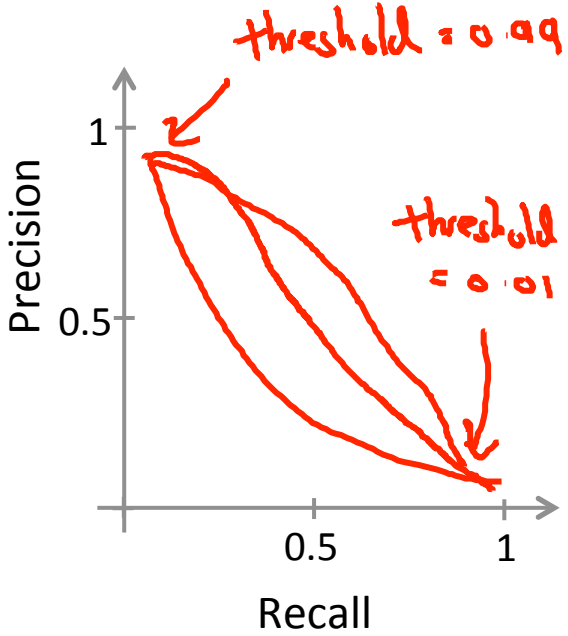
  → Higher precision, lower recall

- Suppose we want to avoid missing too many cases of cancer (avoid false negatives).

  → Higher recall, lower precision.

More generally: Predict 1 if $h_\theta(x) \geq$ threshold.

- precision = $\frac{\text{true positives}}{\text{no. of predicted positive}}$
- recall = $\frac{\text{true positives}}{\text{no. of actual positive}}$



F_1 Score (F score)

How to compare precision/recall numbers?

	Precision(P)	Recall (R)	Average	F1 Score
Algorithm 1	0.5	0.4	0.45	0.44
Algorithm 2	0.7	0.1	0.4	0.175
Algorithm 3	0.02	1.0	0.51	0.0392

Predict y=1 all the time

Average: $\frac{P+R}{2}$

F_1 Score: $2\frac{PR}{P+R}$

P=0 or R=0 $\Rightarrow$ F-score = 0

P=1 and R=1 $\Rightarrow$ F-score = 1