

# Classification

- Build a system that, given some data, can predict a label for that data
  - system built using prior labeled data.

## ★ Evaluation:

- we can count how many times we're correct in predicting = accuracy

$Y_i$  = true class     $\hat{Y}_i$  = predicted class for obs  $i$

$$\text{accuracy} = \frac{1}{N} \sum_{i=1}^N (Y_i == \hat{Y}_i)$$

## Error types:

- for binary classification,

	Predicted +ve	Predicted -ve
Positive eg.	True +ve	False -ve
Negative eg.	False +ve	True -ve

(errors)

## Precision:

"If we say sth is true, how much confidence can we have that it ~~certainly~~ is true?"

- % of things that were classified as +ve & actually were +ve

$$\text{Precision} = \frac{TP}{TP + FP}$$

## Recall:

- percentage of true positives (sensitivity) correctly identified

$$\text{Recall} = \frac{TP}{TP + FN}$$

## f-measure:

- weighted harmonic mean of precision & recall

$$f = (2 * (\text{precision})(\text{recall}) / (\text{precision} + \text{recall}))$$

or we true rate     $\text{TPR} = \frac{TP}{TP + FN}$  (s recall)

also true rate =  $\text{FPR} = \frac{FP}{FP + FN}$