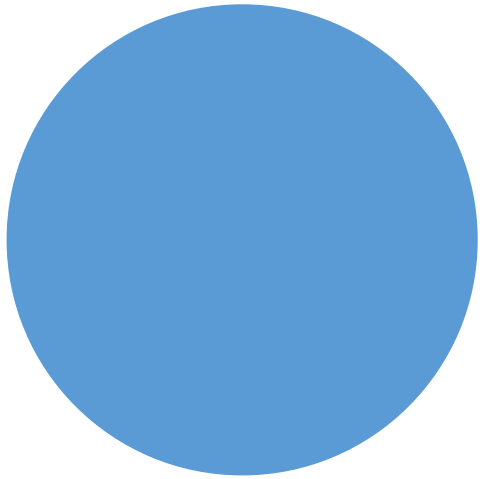
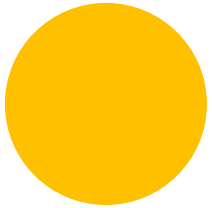



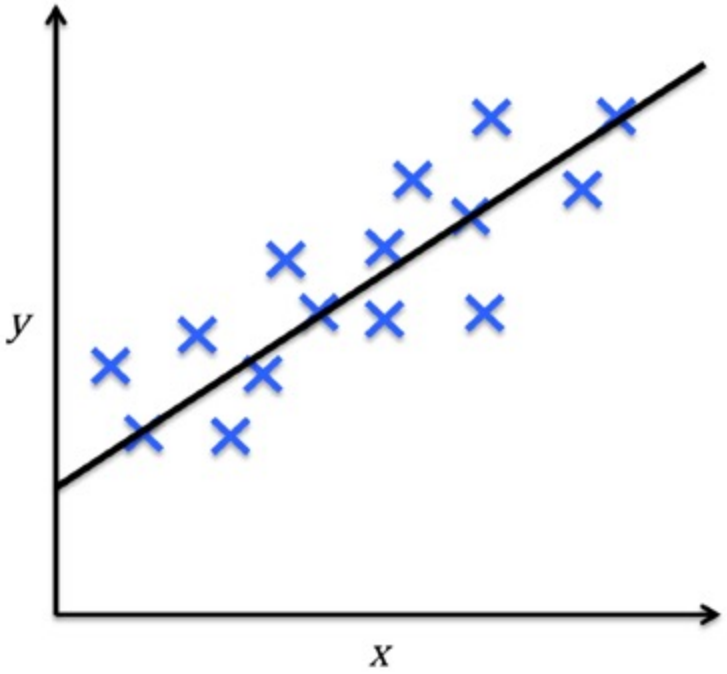
L5: ML evaluation

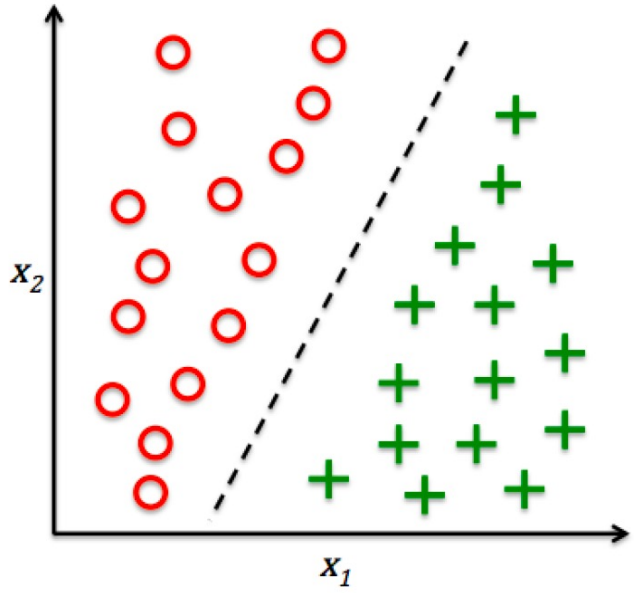
Prof. Xun Jiao

Before class

- HWs uploaded to BB
- Due by next Wed class

Regression and Classification





Regression evaluation

- MAE, MAPE, MSE, RMSE

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y'_i - y_i|$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{y'_i - y_i}{y_i}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Outline

- Classifier evaluation
- Supervised learning process

Confusion Matrix

- A specific table layout that allows visualization of the performance of an algorithm
  - Each row of the matrix represents the instances in a predicted class
  - Each column represents the instances in an actual class
  - For example:

		Actual class	
		Cat	Dog
Predicted class	Cat	5	2
	Dog	3	3

  - How many animals in this dataset? #dogs? #cats?
  - Now, we can calculate different metrics based on this confusion matrix

Confusion matrix

Actual outcome

Predicted outcome	**Predict Sick Actual sick**	**Predict sick Actual not sick**
	Predict Not sick Actual Sick	Predict not sick Actual not sick

		Actual class	
		Cat	**Dog**
Predicted class	**Cat**	**5**	2
	Dog	3	**3**

	Actual -- True/False	
Predicted -- Positive/Negative	True Positive	False Positive (Type I)
	False Negative (Type II)	True Negative

Prediction accuracy

- Percentage of correctly classified examples out of all predictions made.
- PA = correctly_predicted_samples/total_predicted_samples

		Actual class	
		Cat	**Dog**
Predicted class	**Cat**	**5**	2
	Dog	3	**3**

	Actual -- True/False	
Predicted -- Positive/Negative	True Positive	False Positive (Type I)
	False Negative (Type II)	True Negative

- What is the prediction accuracy here?
- How many true positive and true negative?
- PA = (#TP + #TN) / #total_predictions

Precision

- **precision** is the fraction of relevant instances among the retrieved instances

Precision = TP / (TP + FP)
How much what I say is correct

	Actual -- True/False	
Predicted -- Positive/Negative	True Positive	False Positive (Type I)
	False Negative (Type II)	True Negative

Recall

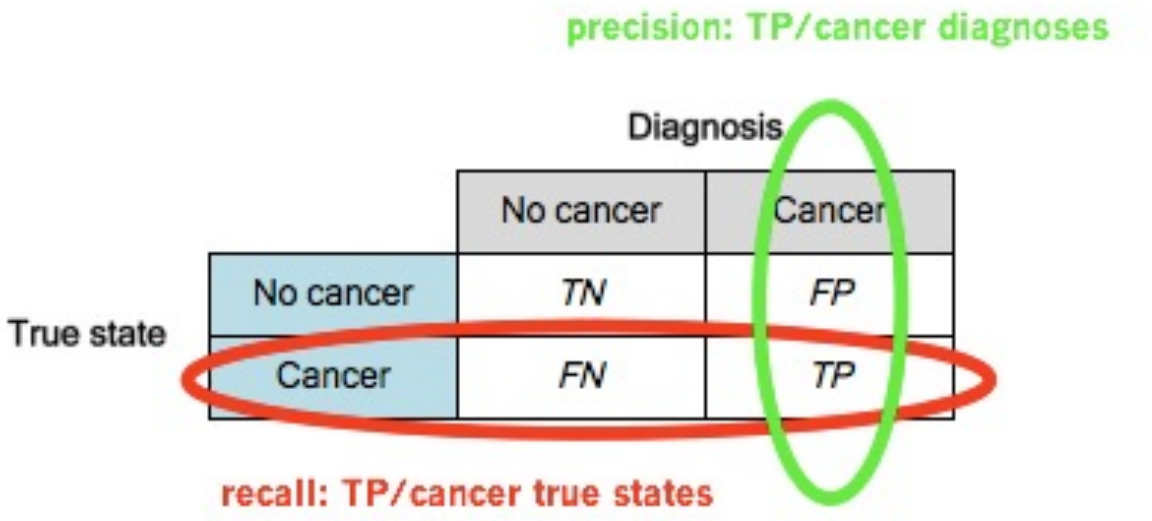
- **recall** (also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.

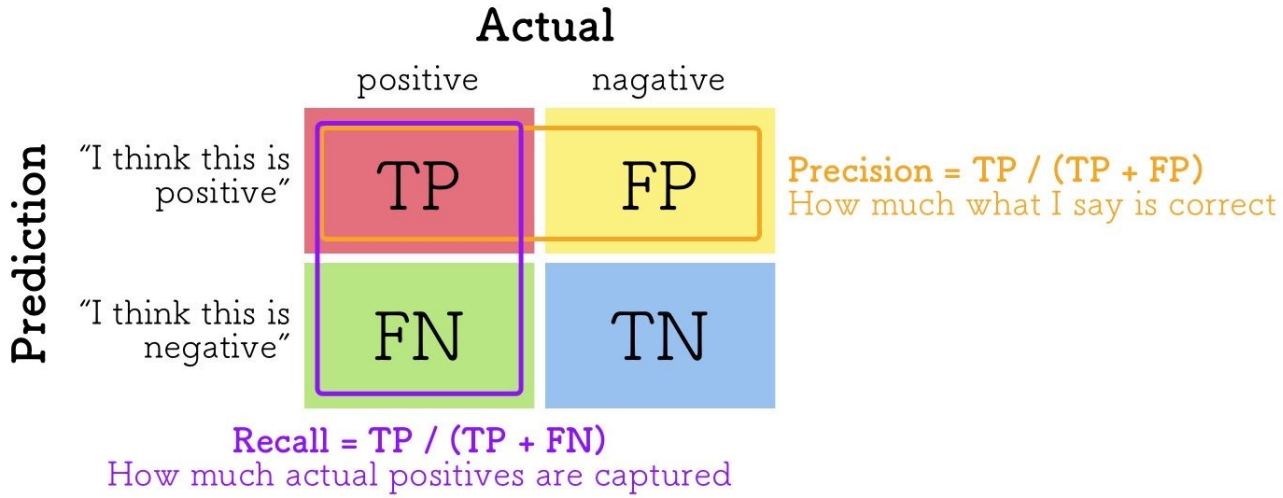
Recall = TP / (TP + FN)
How much actual positives are captured

	Actual -- True/False	
Predicted -- Positive/Negative	True Positive	False Positive (Type I)
	False Negative (Type II)	True Negative

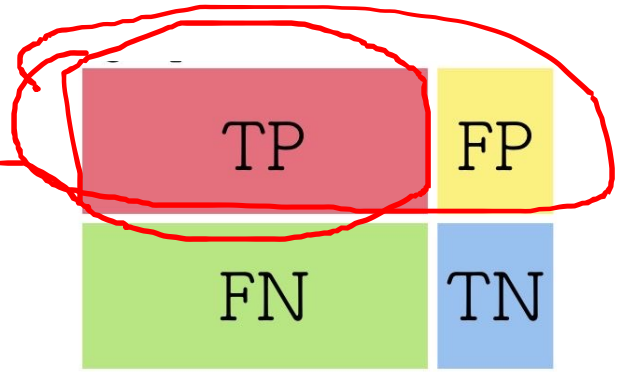
Precision and Recall

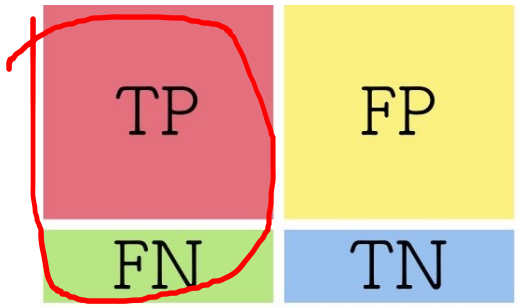
precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while **recall** (also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances.





High precision, low recall?
Low recall, high precision?





F1 score

$$F_1 = \left(\frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} \right) = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

Considers both the precision *p* and the recall *r* of the test to compute the score
Best value? Worst value?

What is the downside of F1 score?

Check yourself

		True/Actual	
		Positive ()	Negative
Predicted	Positive ()	5	1
	Negative	2	2

Accuracy?
Precision?
Recall?
F1 score?