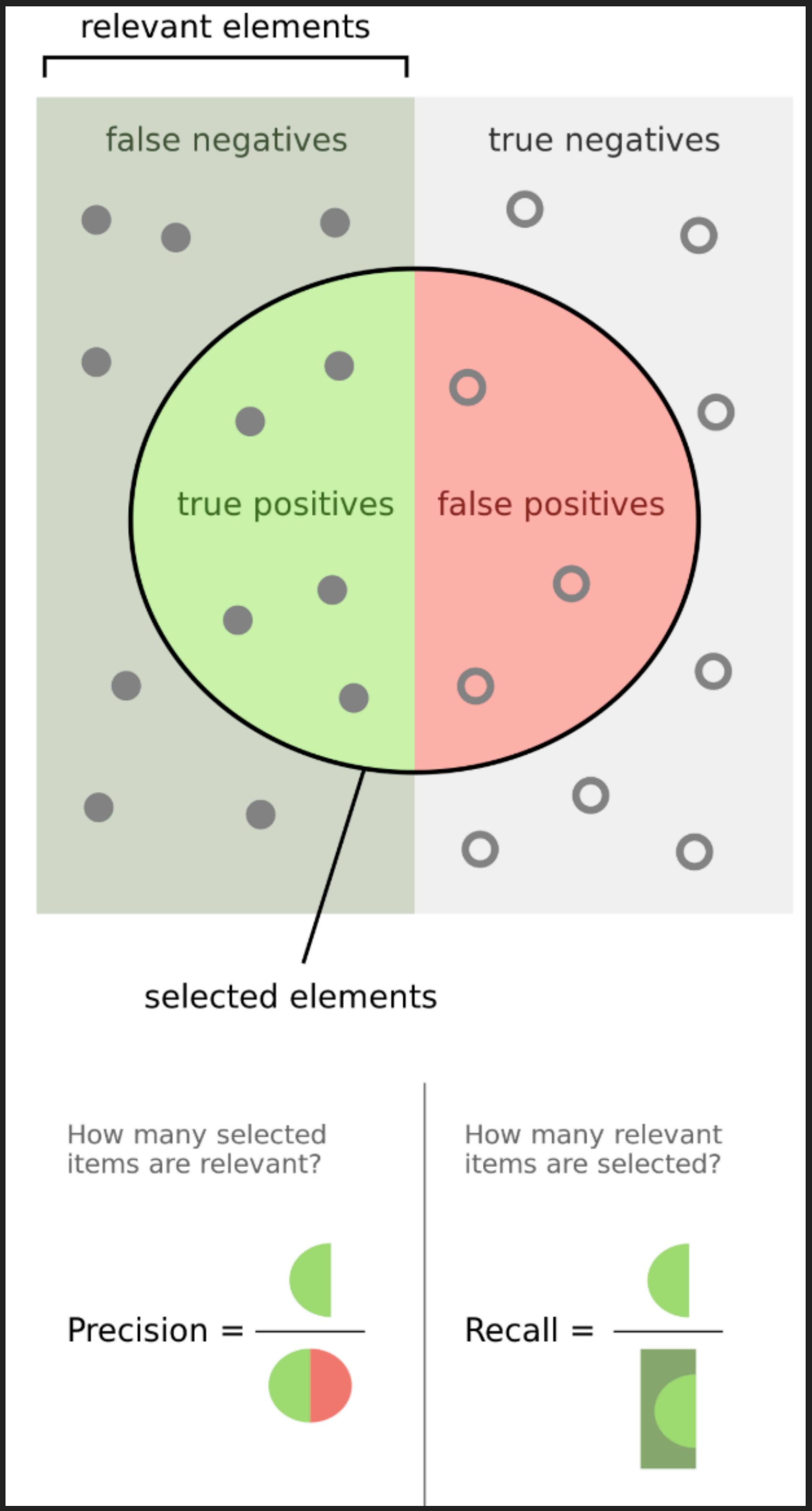
CLASSIFICATION

EVALUATION METRICS

CONFUSION MATRIX

- A confusion matrix is a summary of prediction results on a classification problem
- The number of correct and incorrect predictions are summarized with count values and broken down by each class


relevant elements
false negatives
true negatives
true positives
false positives
selected elements
How many selected items are relevant?
How many relevant items are selected?
Precision =
Recall =

True Positives(TP):

- When we predict positive and the actual is positive

False Positives(FP):

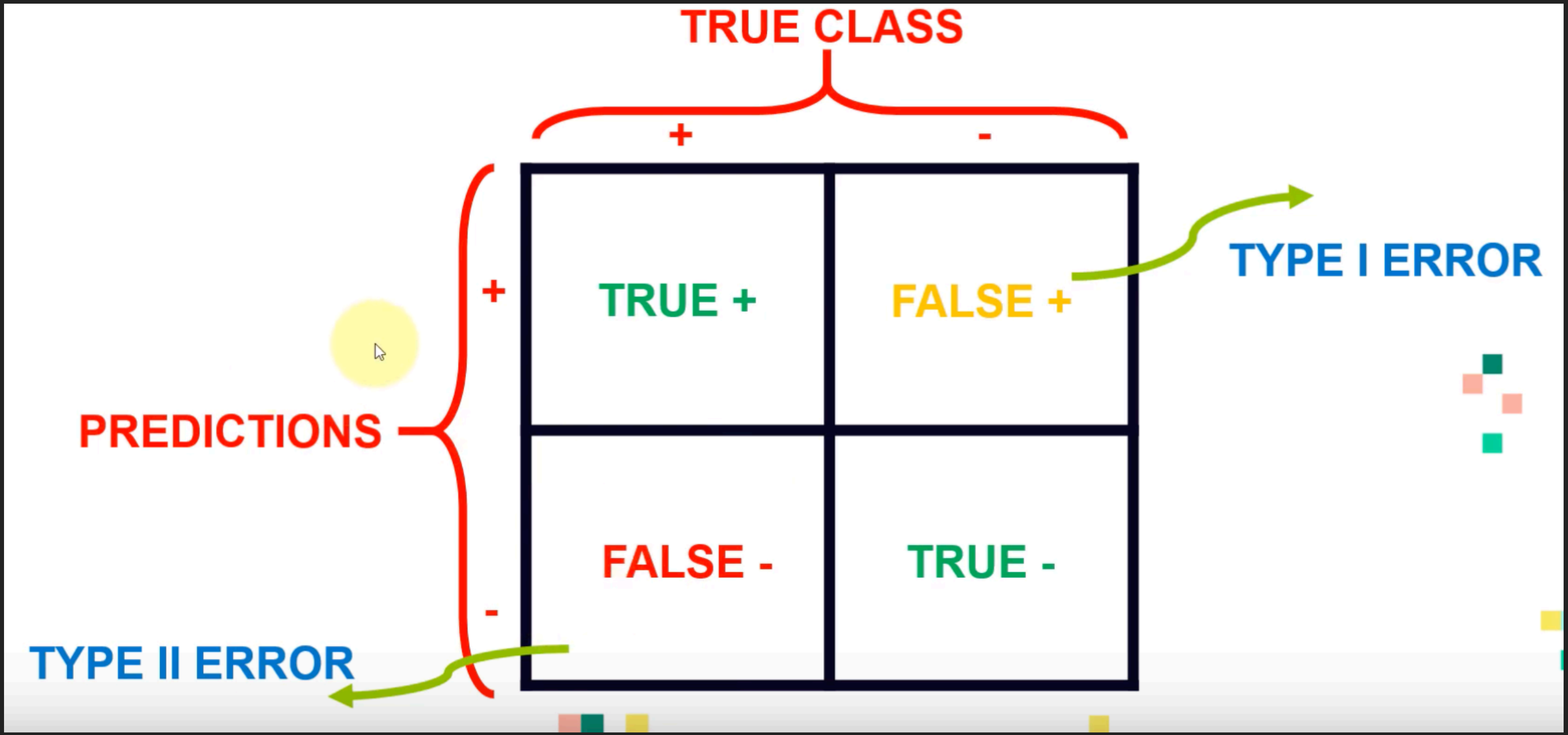
- When we predict that someone is positive and the actual is negative.

False Negatives(FN):

- When we predict that someone is negative and the actual result from the blood test is positive.

True Negatives(TN):

- When we predict that someone is negative and the actual result from the blood test is negative


TRUE CLASS
+
-
PREDICTIONS
+
-
TRUE +
FALSE +
FALSE -
TRUE -
TYPE I ERROR
TYPE II ERROR

Accuracy:

- Overall, how often is the classifier correct?

All correct / total = (TP+TN) / TP+FP+TN+FN

Error Rate / Misclassification Rate:

- Overall, how often is it wrong?

All incorrect / total = (FP+FN) / TP+FP+TN+FN

Specificity (True Negative Rate):

- When it's actually no, how often does it predict no?
- Equivalent to 1 minus False Positive Rate

True negatives/ all negatives (actual no) = TN / TN+FP

Sensitivity or Recall (True Positive Rate):

- When the class was actually positive, how often does it predict positive?

True positives/all positives (actual yes) = TP / TP+FN

Precision:

- When model predicted positive class, how often it correct?

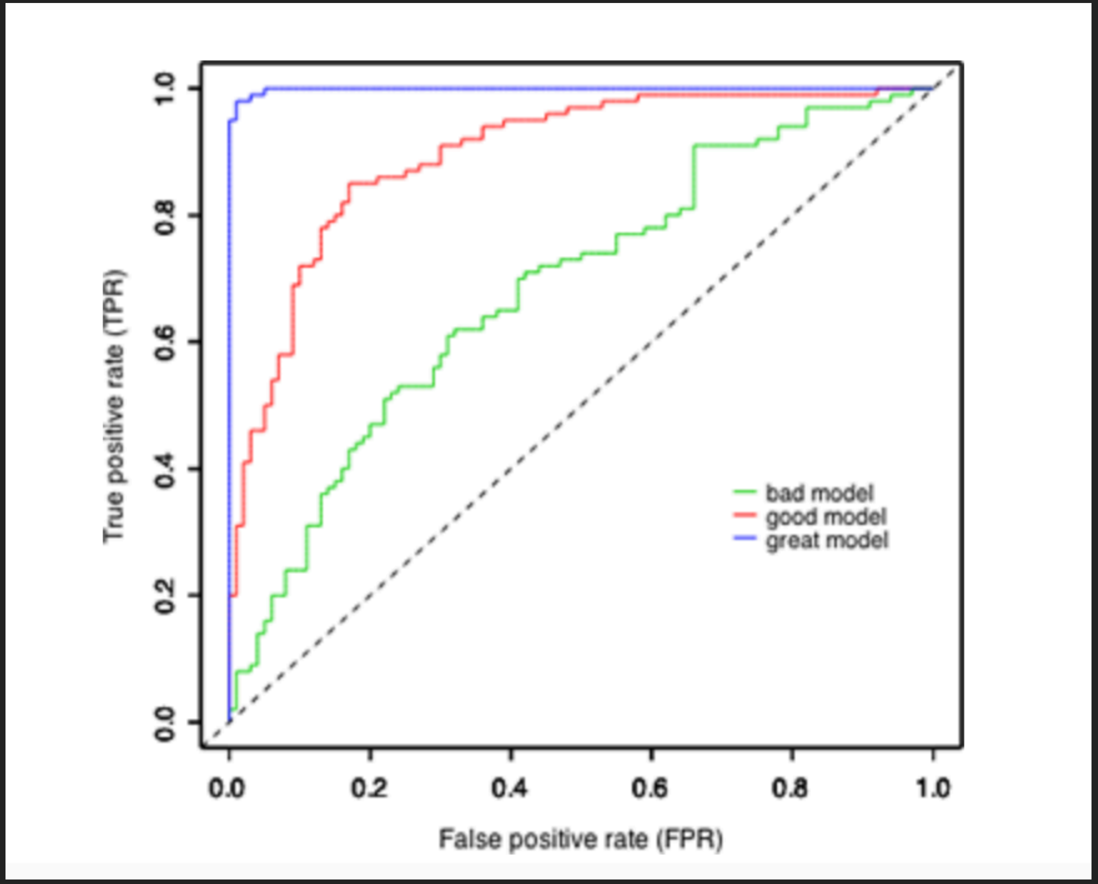
True positives / predicted positives = TP / TP+FP

Other metrics

ROC - AUC score

- AUC score is that if you randomly choose a positive case and a negative case, the probability that the positive case outranks the negative case according to the classifier

- The value can range from 0 to 1. However AUC score of a random classifier for balanced data is 0.5



F1 Score:

This is a weighted average of the true positive rate (recall) and precision

Kappa:

Cohen's kappa statistic is a very good measure that can handle very well both multi-class and imbalanced class problems

A measure of how well the classifier performed as compared to how well it would have performed simply by chance

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

where,

Po is the relative observed agreement

Pe is the hypothetical probability of chance agreement

- Calculation:

		B	
		Yes	No
A	Yes	a	b
	No	c	d

- Po = a+d / a+b+c+d
- P yes = a+b / a+b+c+d
- P no = c+d / a+b+c+d
- Pe = P yes + P no

Python: sklearn.metrics.cohen_kappa_score