Pytorch course
02. Custom metrics

Metrics to evaluate your machine learning models

Generally when deep learning models are calculated, a metric is needed to weight the model. Choosing the right metric **is crucial** while **evaluating** machine learning (**ML**) models.

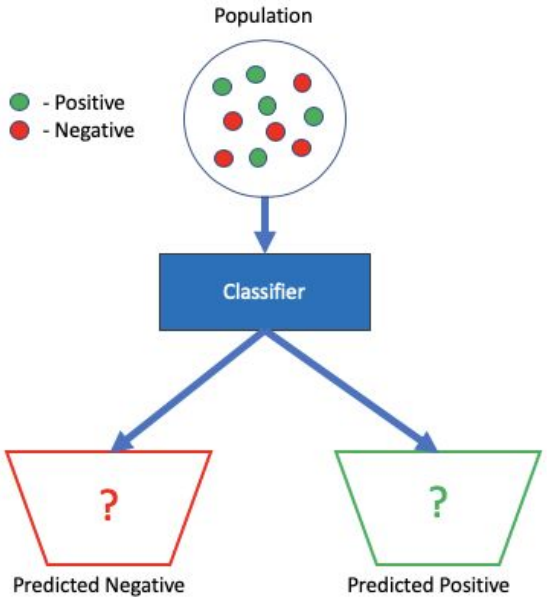
In some applications looking at a single metric may not give you the whole picture of the problem you are solving, and you **may want to use a subset of the metrics** to have a concrete evaluation of your models.

Accuracy, precision, sensitivity, AUC, **recall**, etc…

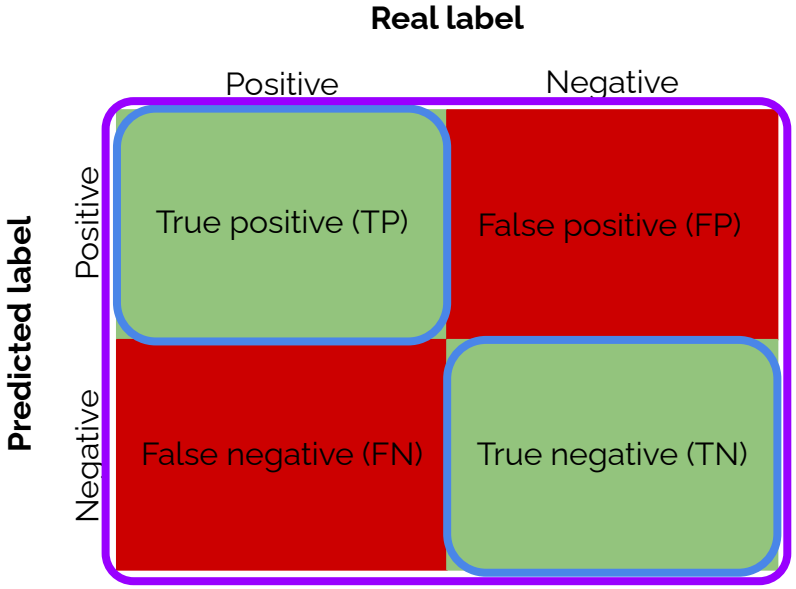
Nevertheless, each of the metrics have a set of true and false positives and negatives samples, that is needed to know what it means

Metrics to evaluate your machine learning models

- ★ **True positive samples (TP):** are **positive outcomes that the model predicted correctly.**
  Example: this means that **patients who were predicted** to have cancer by the model **indeed does have cancer**
- ★ **True negative samples (TN):** are **negative outcomes** that the model **predicted correctly.**
  Example:, this means that **patients who were predicted** to be healthy **indeed does not have cancer.**
- ★ **False positive samples (FP):** are **positive outcomes** that the **model predicted incorrectly.** This is also known as Type I error.
  Example: this means that patients who were predicted to have cancer were actually health.
- ★ **False negative samples (FN):** are **negative outcomes** that the model **predicted incorrectly.** This is also known as Type II error.
  Example: this means **that patients who were predicted** to be healthy **actually had cancer**.



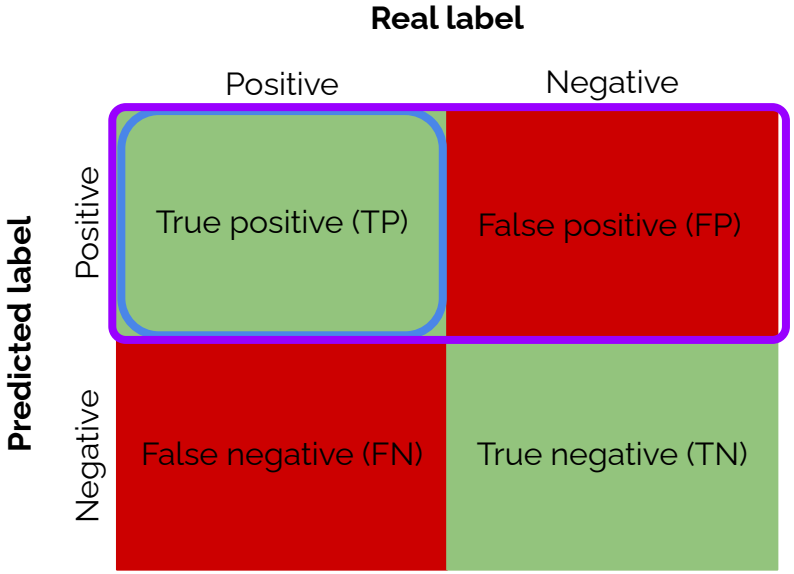
Metrics to evaluate your machine learning models



$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Accuracy is the metric that counts the proportion of correct predictions.

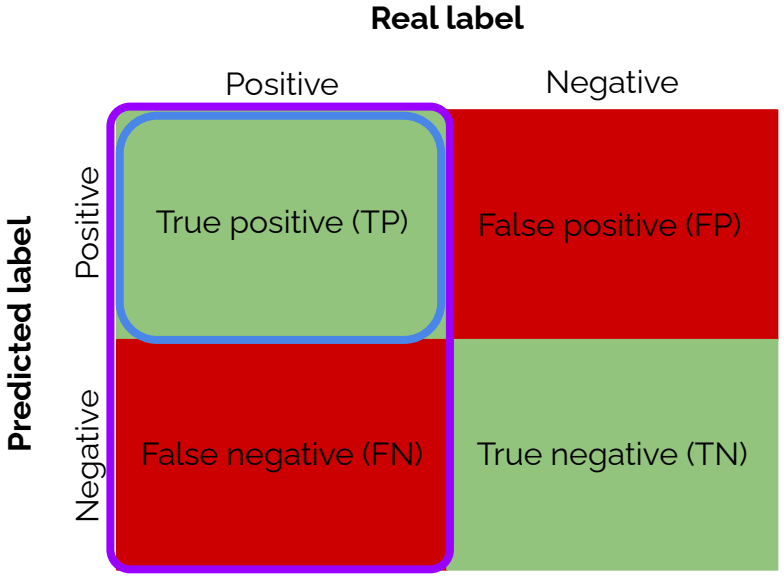
Metrics to evaluate your machine learning models



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

Precision is the metric that questions out of all the examples that predicted as positive, how many are really positive?

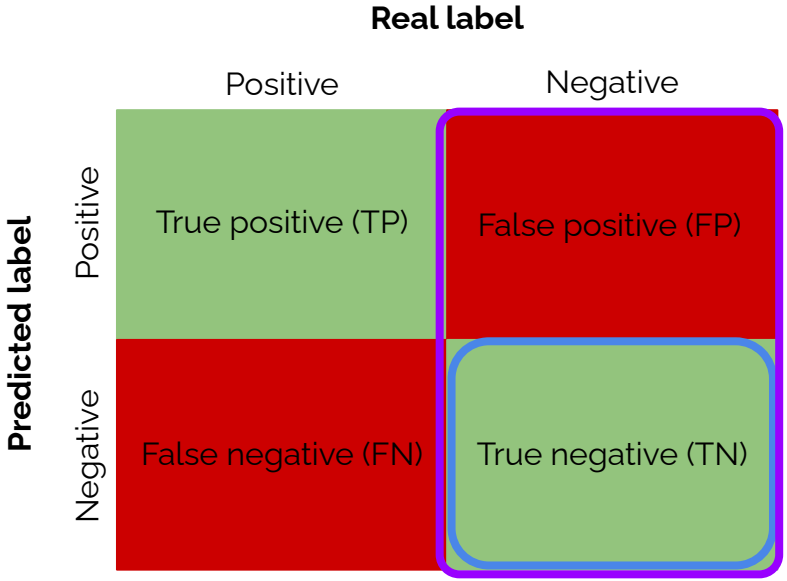
Metrics to evaluate your machine learning models



$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

Sensitivity is the metric that questions out of all the positive examples, how many are predicted as positive? Also known as positive accuracy

Metrics to evaluate your machine learning models



$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

Specificity is the metric that questions out of all the people that do not have the disease, how many got negative results? Also known as negative accuracy