

# HOML - Chapter 3:

## Classification

### Performance Measures

#### Cross Validation

Cross Validation é quando dividimos o nosso training set em  $n$  partes, e utilizamos  $n-1$  partes para treinar o nosso modelo e 1 parte para medir a acuracia do mesmo.

Então repetimos esse processo  $n$  vezes (cada vez com uma combinação diferente), tiramos a média de acuracia dos  $n$  testes e assim teremos uma garantia melhor do nosso resultado

#### Confusion Matrix

A confusion matrix vai nos dizer quantas vezes acertamos uma predição e quantas vezes erramos, dividindo essas predições erradas em subgrupos, cada subgrupo representando uma classe diferente.

Por exemplo, na tabela a seguir:

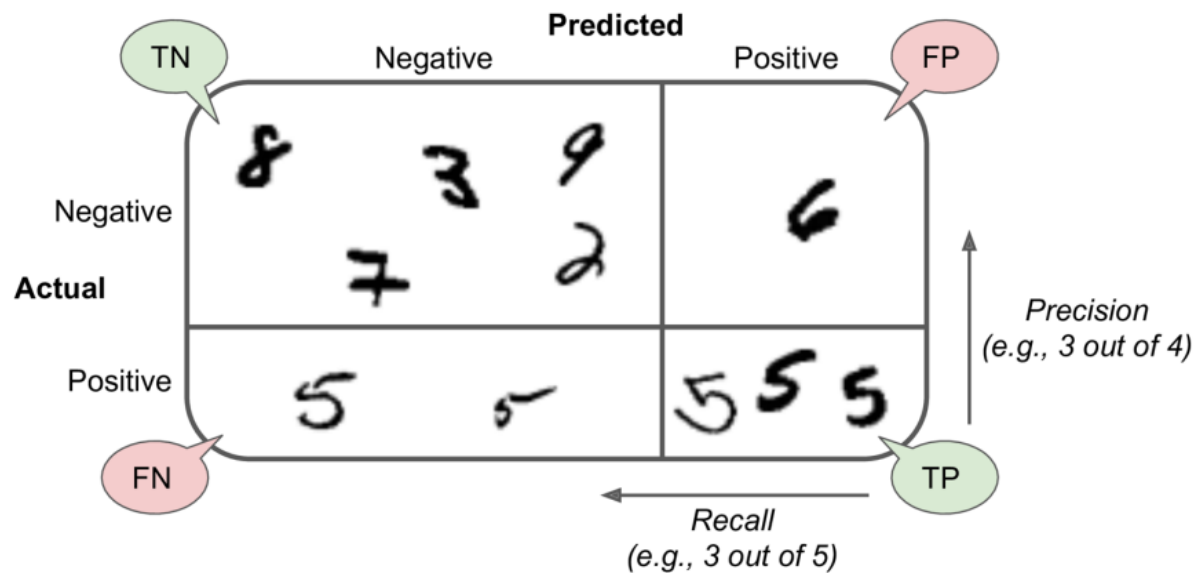
```
array([[53057, 1522],
       [ 1325, 4096]])
```

Temos que para as instâncias que pertencem à primeira classe, 53057 foram previstas como sendo da primeira classe e 1522 da segunda. Das que de fato pertencem à segunda classe, 1325 foram classificados de maneira errônea como na primeira e 4096 classificadas corretamente como da segunda.

Cada linha representa a classe de fato, e cada coluna representa a classe prevista.

## Precision / Recall

Sometimes we may want something more concise than the Confusion Matrix.



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

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

These two metrics are only useful when used together, because usually, extremely high precision will mean a low recall and vice versa