# Evaluation Metrics

Come valutiamo il risultato della nostra predizione?

# Evaluation Metrics

- Cost function $\mathcal{J}$ could be used to quantify a ML algorithm performance. Usually, a performance measure related to the problem semantic is to be preferred
- In a classification problem, the classification accuracy (expressed as percentage) is used a performance measure

$$\text{Accuracy} = \frac{\text{number correctly classified pattern}}{\text{number of patterns}} \times 100\%$$

$$\text{Error} = 100\% - \text{Accuracy}$$

se massima, ho correttamente classificato tutto! L'errore è l'opposto.

  - Might not work well for highly unbalanced class (where Precision/Recall work better)
- In a regression problem, the RMSE (Root Mean Squared Error) is often used

RMSE legato a funzione obiettivo, nelle REGRESSIONI.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_i (y^{(i)} - t^{(i)})^2}$$

# Confusion Matrix

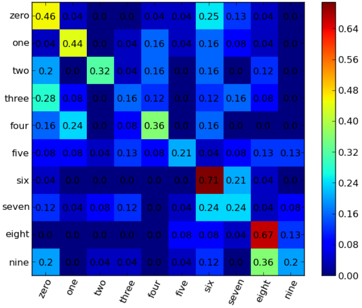
La matrice di confusione grafica la distribuzione degli errori.

- The Confusion Matrix is used in the classification problems to assess the error distribution
- **Example**: Confusion Matrix of a digit classification problem
  - Cell $(r, c)$ contains the fraction of times the algorithm predicted class $c$ when the actual class was $r$.
  - off-diagonal values show errors

46% lo zero è riconosciuto come zero, 25% delle volte è riconosciuto come un '6'

sulla diagonale = 1.00, sugli altri = 0 allora il classificatore è perfetto

true




predicted

# False Positive and False Negative

Casi binari

Given a binary classifier and $T = P + N$ patterns ($P$ positives and $N$ negative instances), we have the possible four cases:

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

- True Positive (TP): a positive pattern has been correctly classified as positive dico "Positivo" e ho predetto bene (True)
- True Negative (TN): a negative pattern has been correctly classified as negative dico "Negativo" e ho predetto bene (True)
- False Positive (FP): a negative pattern has been incorrectly classified as positive. Type I error. dico "Positivo" ma sbaglio (False)
- False Negative (FN): a positive pattern has been incorrectly classified as negative. Type II error. Dico "Negative" ma sbaglio (False)

# Accuracy & Error

Given a binary classifier and $T = P + N$ patterns ($P$ positives and $N$ negative instances)

Non basta essere accurati, perchè bisogna pesare anche gli errori che facciamo! Servono altri indici.

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

$$\text{accuracy} = \frac{TP + TN}{P + N}$$

$$\text{error} = 1 - \frac{TP + TN}{P + N}$$

$$= \frac{FP + FN}{P + N}$$

ciò che classifico correttamente (True davanti)

tutti i positivi e negativi

hanno False davanti

# Precision & Recall

## Precision

- the fraction of positive predictions that are correct
- P(is pos | predicted pos)

quante volte ho detto CORRETTAMENTE positivo
fratto quante volte totale ho detto positivo
(sia predicendo bene che male)

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

Ho predetto 50 volte Positivo, quante volte avevo ragione?

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

## Recall

- the fraction of positive instances that are correctly identified
- P(predicted pos | is pos)

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

In totale ci sono 40 positivi, quanti ne ho indovinati?

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

# Precision & Recall

Da sola è inutile la Precision, perchè se indovino una volta sola,
allora ho massima Precision.
(es: tiro un rigore in carriera, segno. Statistica dice che ho 100% precisione).

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

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

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

| | | Predicted Class | |
|---|---|---|---|
| | | Yes | No |
| Actual Class | Yes | TP | FN |
| | No | FP | TN |

Se dico sempre "positivo", allora quando ho una istanza positiva la prendo sempre, quindi recall 100%.

In a classification task:

- A precision score of 1.0 for a class C means that every item labelled as belonging to class C does indeed belong to class C
  - but says nothing about the number of items from class C that were not labelled correctly) non mi dice quelli sbagliati!
- A recall of 1.0 means that every item from class C was labelled as belonging to class C
  - but says nothing about how many items from other classes were incorrectly also labelled as belonging to class C

Se dico sempre "positivo" è massima la recall

# Precision & Recall: F1 score

- It is often convenient to combine precision and recall into a single metric called the F1 score,
- The F1 score is the harmonic mean of precision and recall

$$F1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

  - harmonic mean gives more weight to low values.
  - The classifier has high F1 score only if both recall and precision are high.