Metriche di Performance

**Matrice di Confusione**

Usata nei problemi di classificazione dove l’output è di 2 o più classi.

* Ogni **riga** si riferisce alle **classi previste** dal classificatore
* Ogni **colonna** per le **classi effettive** registrate nel dataset

Vengono definite 4 categorie:

* **Vero positivo**, tp, Positivi classificati correttamente
* **Vero negativo**, tn, Negativi classificati correttamente
* **Falso positivo**, fp, Negativi, ma classificati come positivi
* **Falso negativo**, fn, Positivi, ma classificati come negativi

Obiettivo: massimizzare la diagonale maggiore e minimizzare quella minore. E la soluzione ottimale avrebbe una diagonale minore nulla.

**Accuracy**

Proporzione di istanze classificate correttamente.

*Acc = (Tp + Tn) / Totale*

Buona quando il dataset è bilanciato, quando non lo è, è pessima.

**Error Rate**

Proporzione di istanze classificate male.

*Error rate = 1 - Accuracy*

**True Positive Rate / False Positive Rate**

La proporzione di esempi classificati come positivi (negativi) tra quelli che sono effettivamente veri (falsi).

*TP Rate = Tp / (Tp + Fn)*

*FP Rate = Fp / (Fp + Tn)*

**Precision**

Proporzione di esempi predetti veri che sono effettivamente veri.

*Prec = Tp / (Tp + Fp)*

Buona quando si vogliono minimizzare i falsi positivi.

Quando la classe di minoranza è quella di interesse, Accuracy non è da usare, ma si usa la Precision.

**Recall**

Come il TP Rate: La proporzione di esempi classificati come positivi tra quelli che sono effettivamente veri.

*Recall = Tp / (Tp + Fn)*

Buona per minimizzare i falsi negativi.

Quindi per

* Minimizzare i Falsi negativi, vogliamo una recall alta, senza che la precision sia troppo negativa
* Minimizzare i Falsi positivi, vogliamo una precision alta.

**F1 Score**

Si considera la precision e la recall.

*F1Score = 2PR / (P + R)*

Quando la differenza tra i due valori aumenta, F1 score restituisce un valore che tende al valore più basso dei due.

**ROC Space**

Traccia il tasso dei Tp rispetto al tasso dei Fp. Illustra le prestazioni tra benefici (veri positivi) e costi (falsi positivi).

Un classificatore discreto produce solo un’etichetta di classe. E quindi una coppia (fp, tp) corrispondente a un punto nello spazio ROC.

* (0,1): Classificatore perfetto
* y = x: Prestazioni casuali. Random Guess (scelta casuale della classe)
* Area sotto diagonale: Comportamento peggiore rispetto a quello casuale
* Area alta sinistra: Classificatori buoni, con alto tp e basso fp

Qualsiasi curva generata da un insieme finito di istanze è una funzione gradino. Si dovrebbero prendere un numero molto elevato di soglie.

**AUC ROC**

Probabilità che un classificatore classifichi un’istanza positiva scelta a caso più alta di una negativa scelta a caso. Varia tra [0, 1]

**PR curve**

Traccia la recall (x) e la precision (y). L'osservazione delle curve PR può mettere in luce differenze tra gli algoritmi che non sono evidenti nello spazio ROC.

L’obiettivo è posizionarsi in alto a destra.

Quando i negativi superano di molto i positivi:

* Una grande variazione del numero di Fp può portare a una piccola variazione del FPR utilizzato nell'analisi ROC.
* La Precision confronta i Fp con i Tp, cogliendo l'effetto dell'elevato numero di esempi negativi sulle prestazioni dell'algoritmo.

**Classificatori multiclasse**

La matrice di confusione diventa n\*n con n classificazioni corrette e n2-n possibili errori.

Può essere usato:

* Micro media: Traccia una curva pr-roc per ogni classe, sommando globalmente i valori Tp, Fp, Fn per tutte le classi.
  + Esempio: Tp (quelle predette bene), Fp (predette male, partendo da “predicted”), Fn (predette male, partendo da “label”).
* Macro media: Traccia una curva pr-roc per ciascuna classe, considerandola separatamente.
* No average: Considera le curve senza combinarle.

Metriche di Regressione

* **MSE Errore quadratico medio**: Quando le previsioni si allontanano dai valori reali usando la distanza euclidea. Quanto i dati sono concentrati intorno la linea di miglior adattamento. Quadratico: così gli errori più grandi vengono considerati di più. Sempre non negativo, più piccolo è meglio.
* **RMSE Root MSE**: Deviazione standard dei residui. Sempre non negativo, più piccolo è meglio.
* **MAE Errore assoluto medio**: Media dei valori assoluti della differenza tra originale e valore previsto su tutte le istanze. Sempre non negativo, più piccolo è meglio.

MSE è differenziabile, gli errori più elevati pesano di più e le unità di misura sono al quadrato. RMSE riporta l’errore all’unità corretta, mantenendo la penalizzazione degli errori più elevati. RMSE è quello migliore da usare.