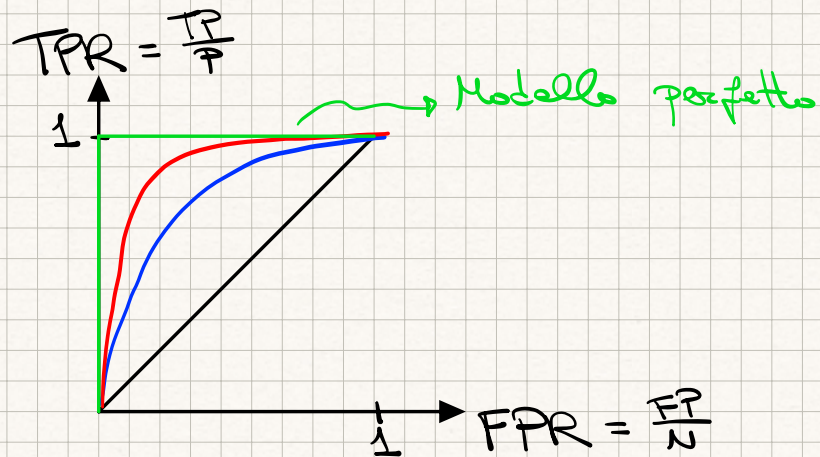


ROC: Receiver Operating Characteristics

È un modo grafico per paragonare la bontà di due classificatori.

Mostra il trade-off presente fra i True Positive Rate e False Negative Rate

L'area sotto la ROC curve è l'accuratezza del modello.



Avremo True Positive Rate in y e False Positive Rate in x.

In termini di curva ottimale abbiamo che $TPR=1$ ovvero che tutte le tuple sono classificate correttamente, mentre $FPR=0$, perché non abbiamo falsi positivi se tutte le tuple vengono classificate correttamente. Da qui possiamo comprendere il classificatore ottimo.

Più la curva si avvicina alla diagonale, meno accurato è il modello.

Per valutare due classificatori dobbiamo semplicemente valutare l'area sotto la curva ROC.

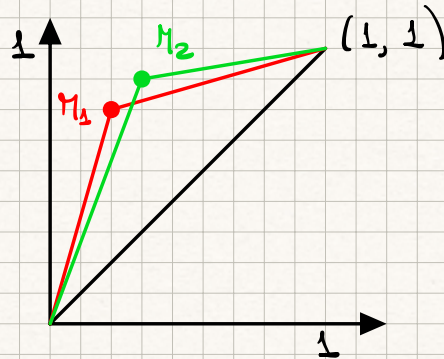
Tipicamente, valutando TPR e FPR ho difatti un solo punto:

(FPR, TPR)

Per ottenere tutti i punti (FPR, TPR) della curva ROC devo usare il modello per classificare il test set, ogni volta con parametri diversi, così da ottenere punti diversi e quindi una curva.

Ma se abbiamo un solo classificatore possibile?

Solitamente non si calcolano diversi punti (FPR, TPR) usando diversi parametri di un classificatore, ma si usa il singolo modello, calcoliamo il punto (FPR, TPR):



Più vicino è il punto al punto ottimo $(0, 1)$ più il classificatore è migliore.

Usando la Cross-Validation possiamo avere una distribuzione di punti a cui applicare un test probabilistico (t-TEST, Wilcoxon Test)

NOTA

Questo è un metodo per valutare classificatori con classi sbilanciate, perché la sola accuratezza non è sufficiente. Se abbiamo un classificatore che lavora molto bene con la classe maggioritaria ma non con quella minoritaria, l'accuratezza non lo noterà.

$$\text{Accuracy} = \frac{TP + TN}{\text{All}}$$

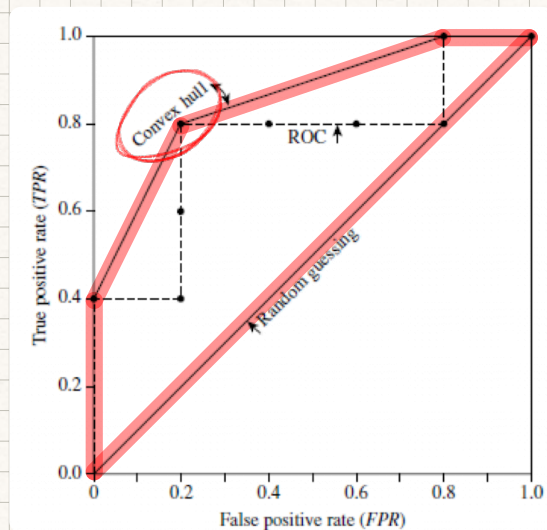
se una delle due classi è sbilanciata allora $TP \gg TN$ o viceversa, dando il bias all'accuracy. Per questo si usano misure come F-Score o la curva ROC.

Esempio Classificatori Probabilistici

Abbiamo detto che per avere più punti su cui creare la ROC curve possiamo semplicemente cambiare dei parametri del classificatore. Nei classificatori probabilistici il parametro da settare è la soglia probabilistica per cui possiamo considerare un'istanza di una classe o dell'altra.

Tuple #	Class	Prob.	TP	FP	TN	FN	TPR	FPR
1	P	0.90	1	0	5	4	0.2	0
2	P	0.80	2	0	5	3	0.4	0
3	N	0.70	2	1	4	3	0.4	0.2
4	P	0.60	3	1	4	2	0.6	0.2
5	P	0.55	4	1	4	1	0.8	0.2
6	N	0.54	4	2	3	1	0.8	0.4
7	N	0.53	4	3	2	1	0.8	0.6
8	N	0.51	4	4	1	1	0.8	0.8
9	P	0.50	5	4	0	1	1.0	0.8
10	N	0.40	5	5	0	0	1.0	1.0

↳ Parametri



Chiamiamo **Convex Hull** il minimo set convesso che contiene la curva ROC

COSTO DI UN CLASSIFICATORE

Il costo di un classificatore può essere rappresentato con il punto (FPR, TPR) , ma è definito come:

$$\text{Cost} = FPR \cdot P(n) c(Y, n) + FPR P(p) c(N, p)$$

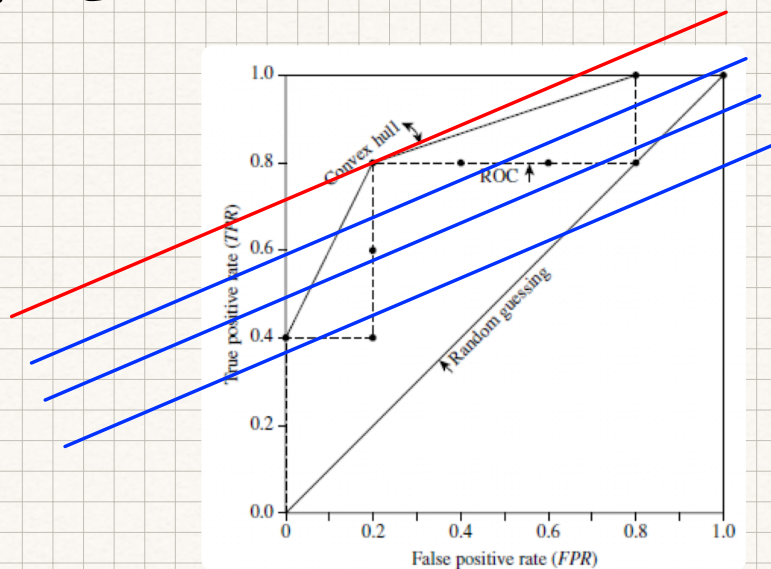
dove $P(n)$ e $P(p)$ è la probabilità di avere la classe negativa o positiva **a priori** rispettivamente per un esempio, mentre $c(Y, n)$ e $c(N, p)$ è il costo del misclassificare un esempio positivamente quando in realtà appartiene alla classe negativa e viceversa per $c(N, p)$.

Cosa vuol dire costo? Se per esempio consideriamo un classificatore per il tumore, il costo $c(N, p)$ è molto alto, perché classifichiamo una persona affetta da tumore come non affetta!

Una volta fissati i valori di $P(n)$, $c(Y, n)$, $P(p)$, $c(N, p)$ possiamo ottenere una famiglia di rette, chiamate **iso-cost lines** che hanno pendenza h :

$$h = \frac{P(n) c(Y, n)}{P(p) c(N, p)}$$

Questa famiglia di rette parallele sono così fatte



Dobbiamo prendere i costi che corrispondono al classificatore migliore. Ovvero bisogna prendere la retta tangente al Convex Hull, ovvero della curva ROC.

Il miglior classificatore, quindi dipende dal costo del missclassificazione.

Il punto è fissare i costi $c(Y, n)$, $c(N, p)$ che tipicamente sono associati al dominio di classificazione. (Medico, legale, ecc...)