

Valutazione delle Prestazioni di un Classificatore

Performance Evaluation

Valutazione delle Prestazioni

- Una volta appreso un classificatore è di fondamentale importanza valutarne le prestazioni
- La valutazione delle prestazioni richiede un attento esame del problema e delle componenti per cui ha senso effettuare una valutazione
- E' necessario utilizzare in modo ragionato le misure di valutazione disponibili
 - Giustificare perché utilizzare una misura piuttosto che un'altra o perché complementare una misura con un'altra al fine di avere un quadro completo delle prestazioni

Modalità di test di un classificatore

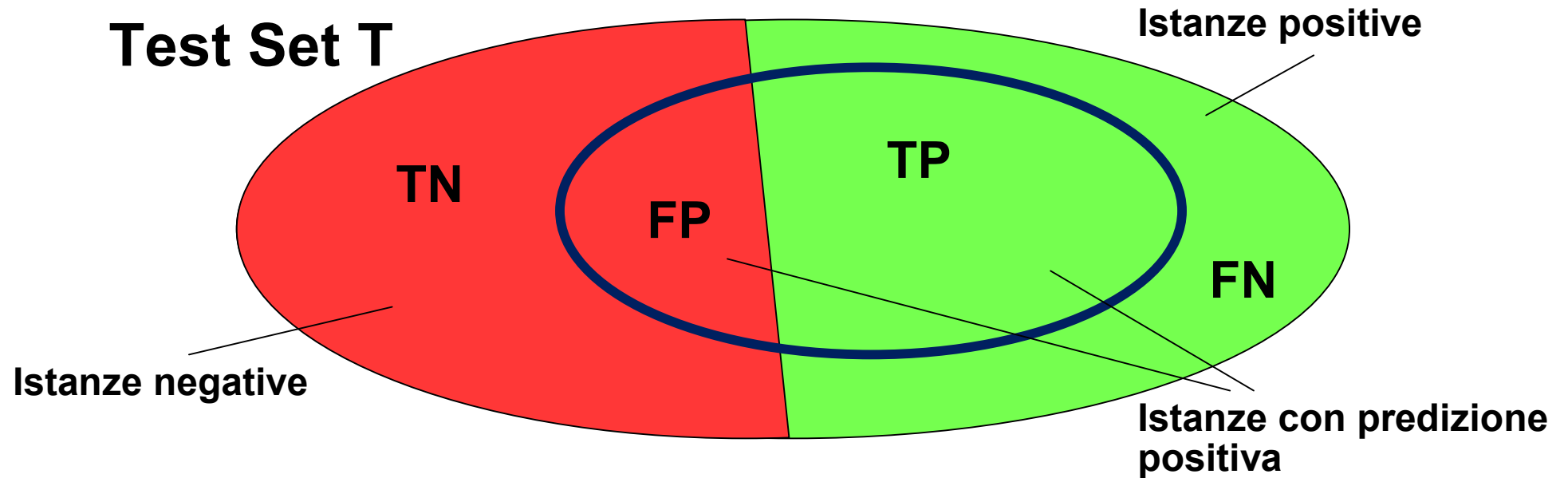
- Sul **training set** stesso
 - overfitting!
- Su un **test set disgiunto** dal training set
- Mediante **cross validation**
 - Suddivido il training set in k sottoinsiemi (**k-fold**)
 - Testo il classificatore k volte allenandolo su $k-1$ sottoinsiemi e testandolo sul sottoinsieme rimanente
 - Le prestazioni complessive possono essere calcolate mediando sui k esperimenti
 - $k = 10$ risulta sperimentalmente un buon numero di fold
- Mediante **split**
 - Suddividiamo l'insieme di dati disponibile in training e test set (normalmente il primo è più grande del secondo, es. 66% e 34%)

Matrice di Confusione

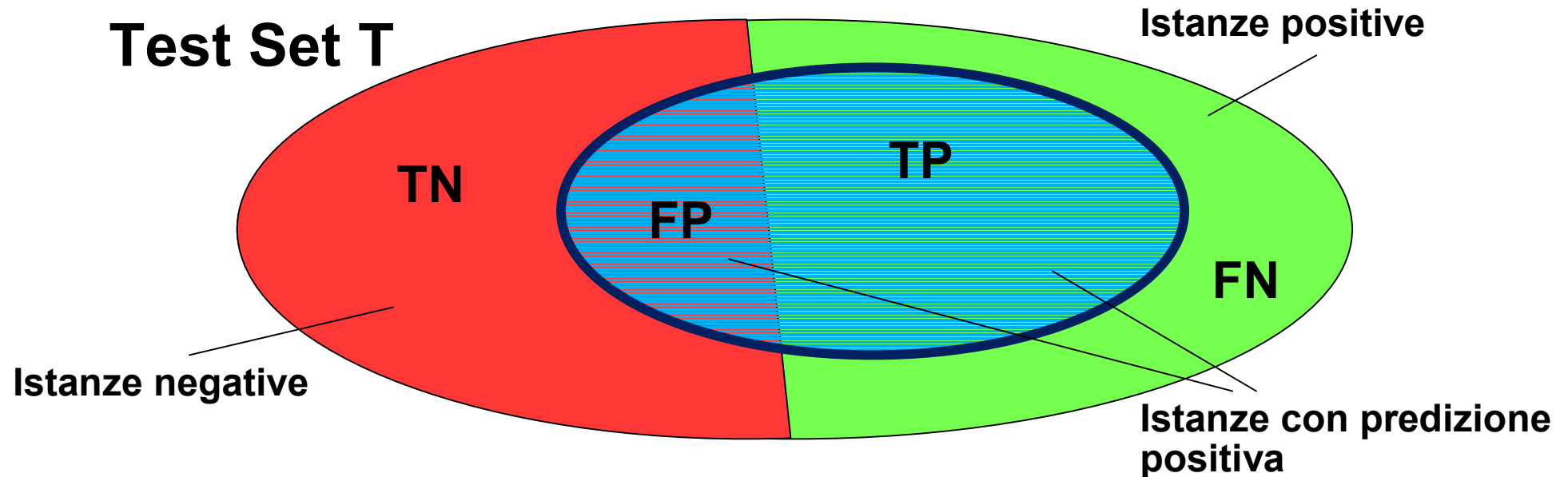
- Applicando il classificatore al test set possiamo visualizzare la distribuzione delle istanze rispetto alla classificazione predetta e a quella reale
- Utilizziamo la cosiddetta **matrice di confusione**
- Es. se la classificazione è binaria:
 - **TP** = true positive
 - **TN** = true negative
 - **FN** = false negative
 - **FP** = false positive

		Classe predetta	
		Positivo	Negativo
Classe reale	Positivo	TP	FN
	Negativo	FP	TN

Misure per valutare le prestazioni (1)



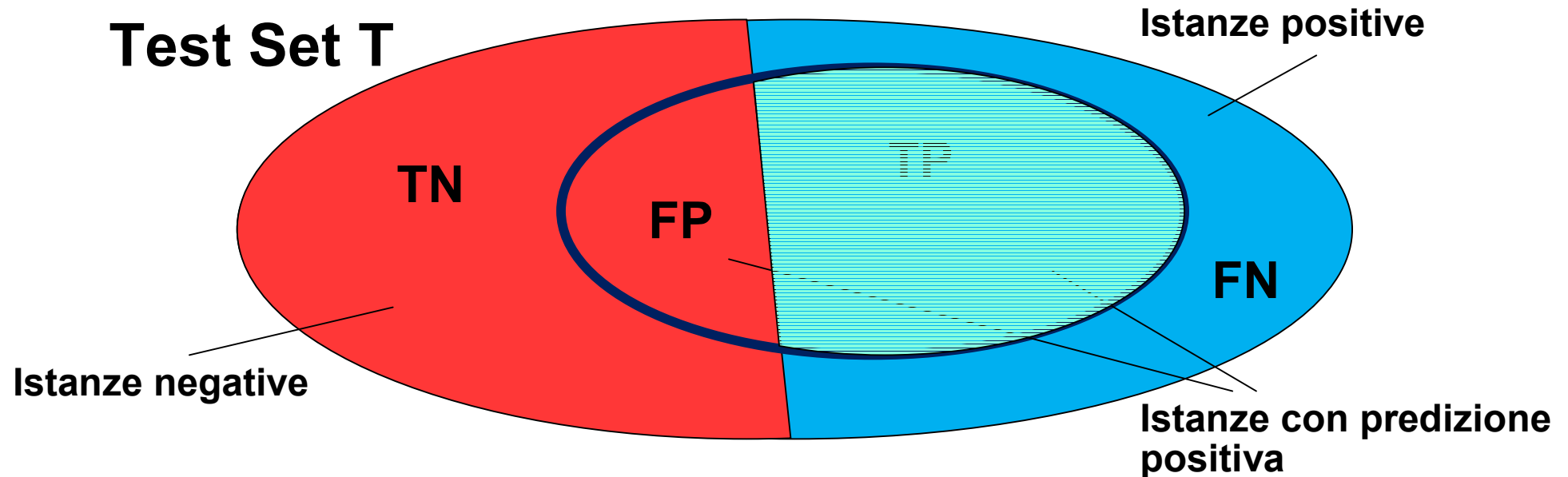
Misure per valutare le prestazioni (2)



• Precisione

- $P = TP / (TP + FP)$
- Percentuale di predizioni positive corrette

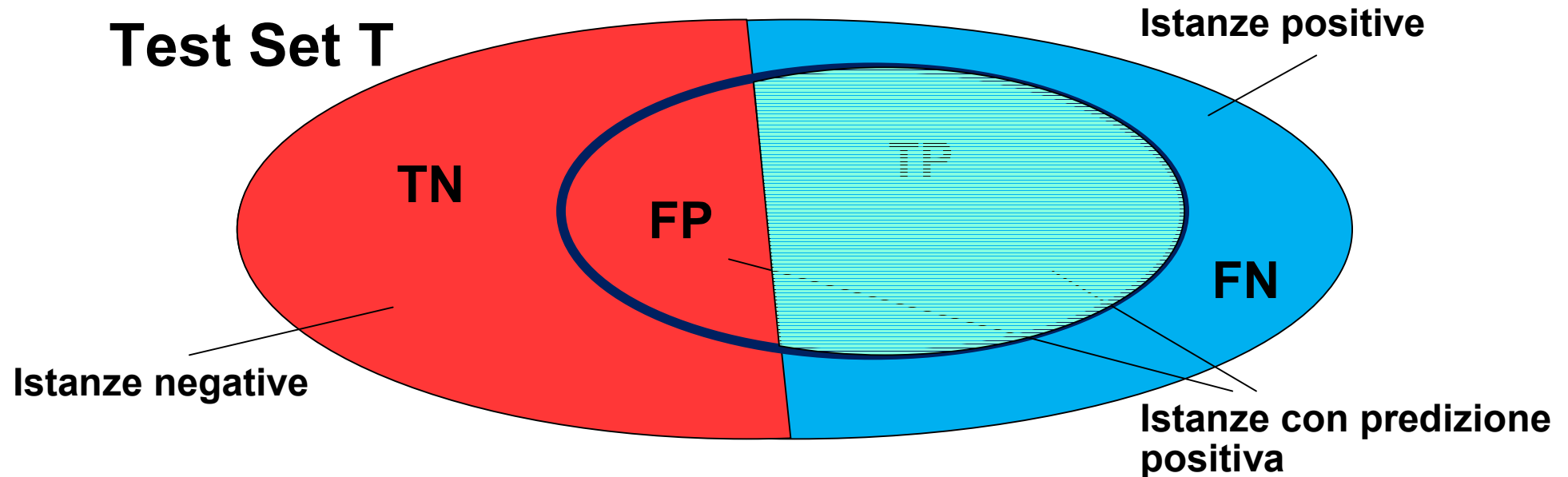
Misure per valutare le prestazioni (3)



• Recall

- $R = TP / (TP + FN)$
- Percentuale di istanze realmente positive classificate come positive

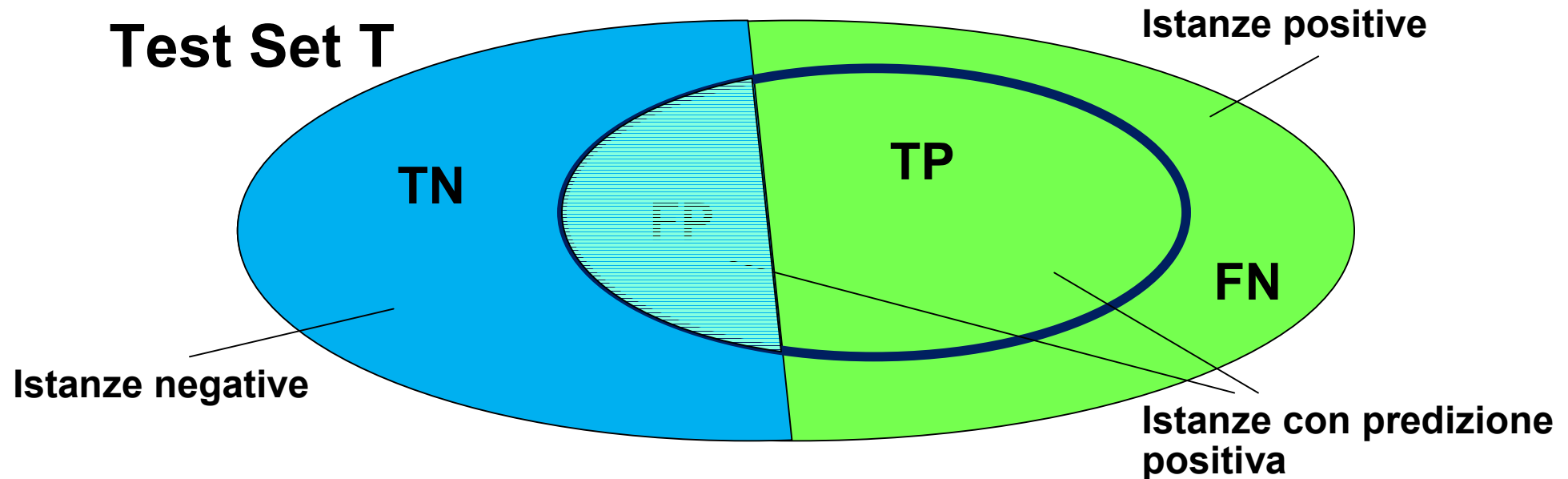
Misure per valutare le prestazioni (4)



- **TP rate = $TP/(TP+FN)$ = Recall**

- Percentuale di istanze realmente positive classificate come positive

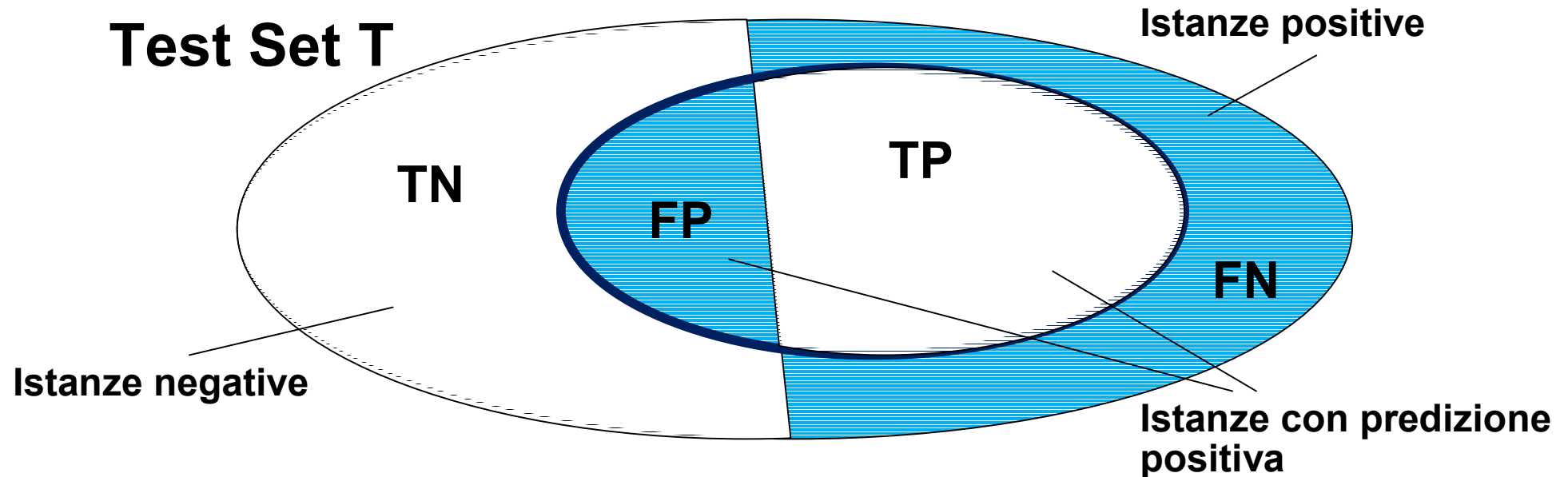
Misure per valutare le prestazioni (5)



- **FP rate = $FP / (FP + TN)$**

- Percentuale di istanze realmente negative classificate erroneamente come positive

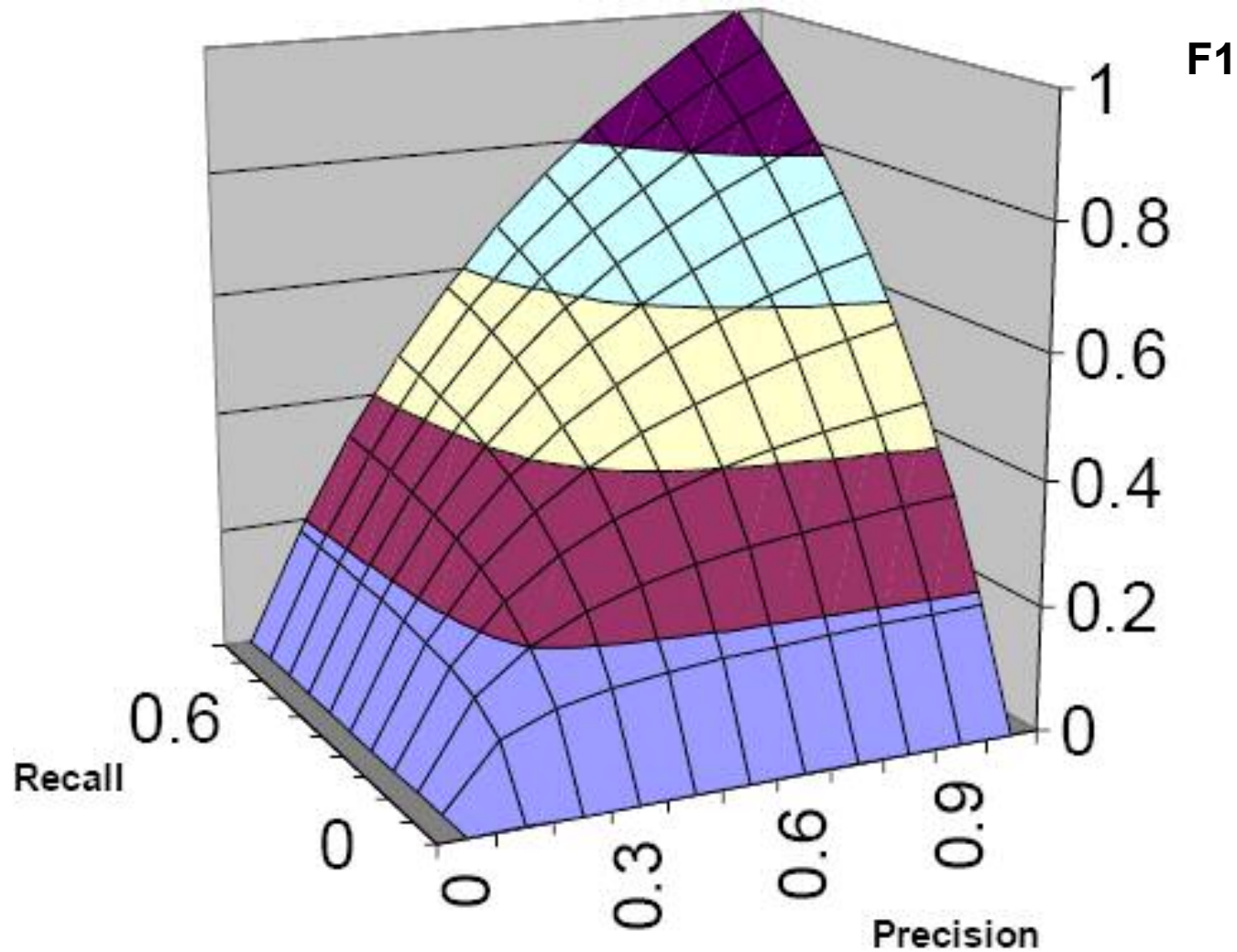
Misure per valutare le prestazioni (6)



- **Accuratezza** = $(TP+TN)/(TP+TN+FP+FN)$
 - Percentuale di istanze sia positive sia negative classificate correttamente

F1 measure

- $F1 = 2PR / (P+R)$
- E' una media armonica tra precisione e recall

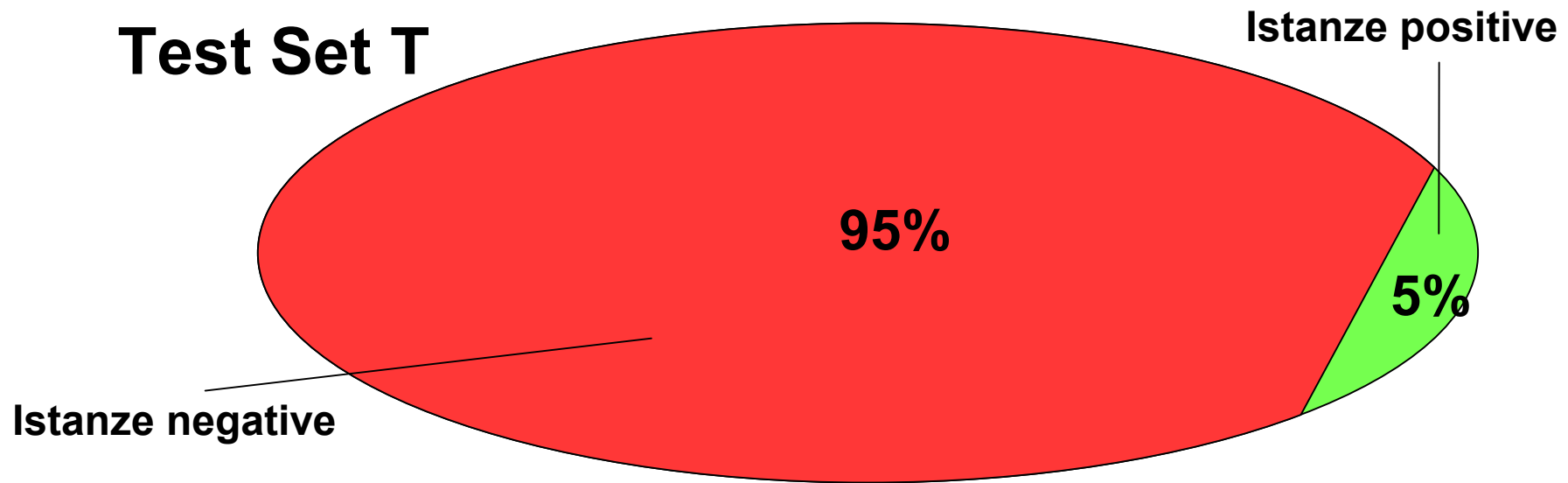


Come utilizzare le misure

- Quando utilizzare la **accuracy**
 - Quando ho un problema multiclasse e tutte le classi sono di interesse
 - Quando ho un problema binario, le due classi sono bilanciate e interessanti
- Quando utilizzare **P**, **R** e **F1**
 - Se c'è almeno una classe non interessante
 - Se voglio studiare l'andamento del classificatore su una classe in particolare

Esempio 1

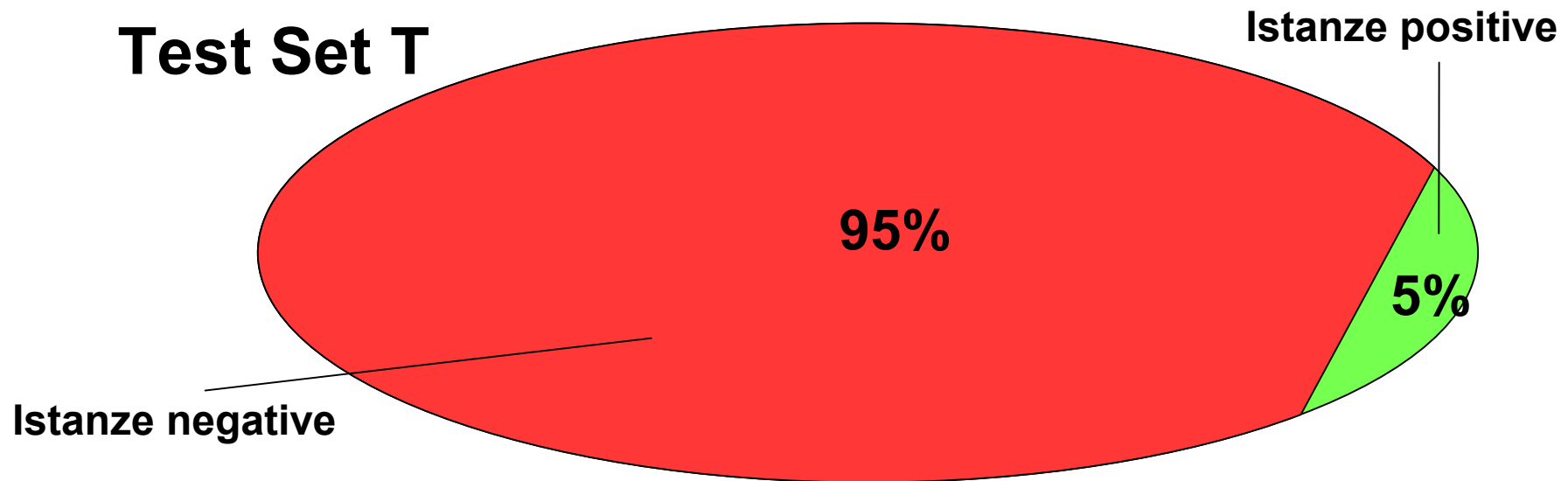
- Classificazione binaria
 - Esempio: il classificatore emette sempre classificazione negativa



- Accuracy = 0.95
- P = 0
- R = 0

Esempio 2

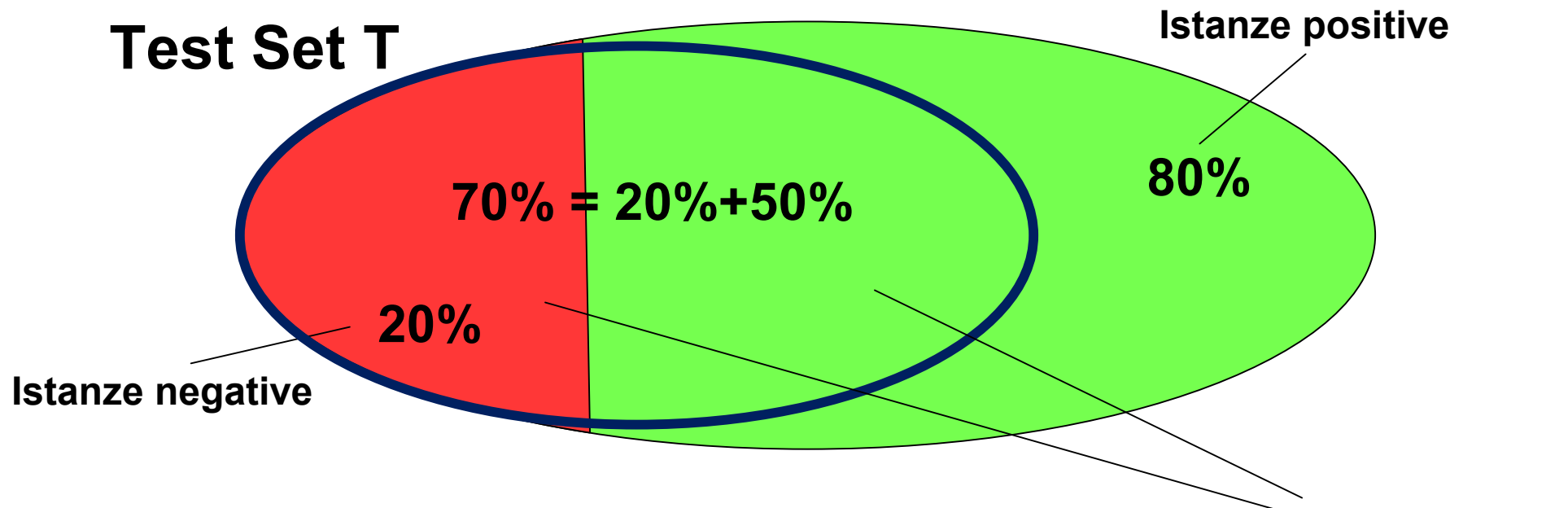
- Classificazione binaria
 - Esempio: il classificatore emette sempre classificazione positiva



- Accuracy = 0.05
- P = 0.05
- R = 1

Esempio 3

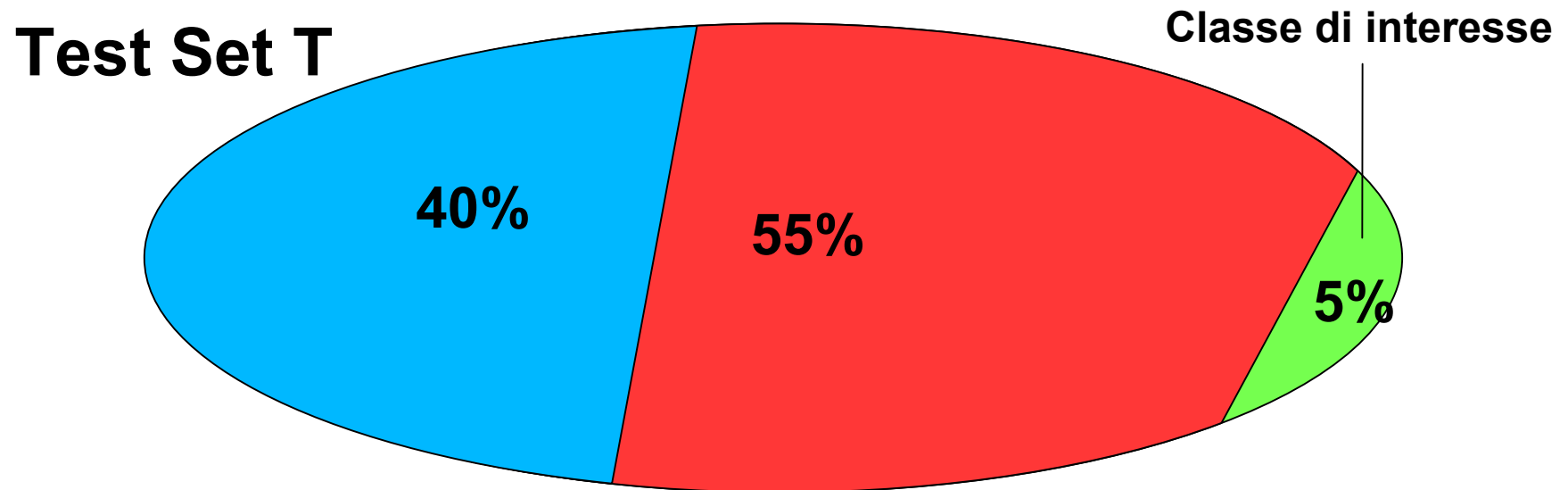
- Classificazione binaria



- Accuracy = $50/100 = 0.5$
- P = $50/70 = 0.71$
- R = $50/80 = 0.62$

Esempio 4

- Classificazione n-aria



- Caso: Il classificatore classifica correttamente tutta la classe di interesse, ma assegna classe rossa agli azzurri e azzurra ai rossi
- Accuracy = $5/100 = 0.05!!!$
- P (rispetto alla classe verde) = $5/5 = 1$
- R (rispetto alla classe verde) = $5/5 = 1$