

Fondamenti di Machine Learning

La Classificazione

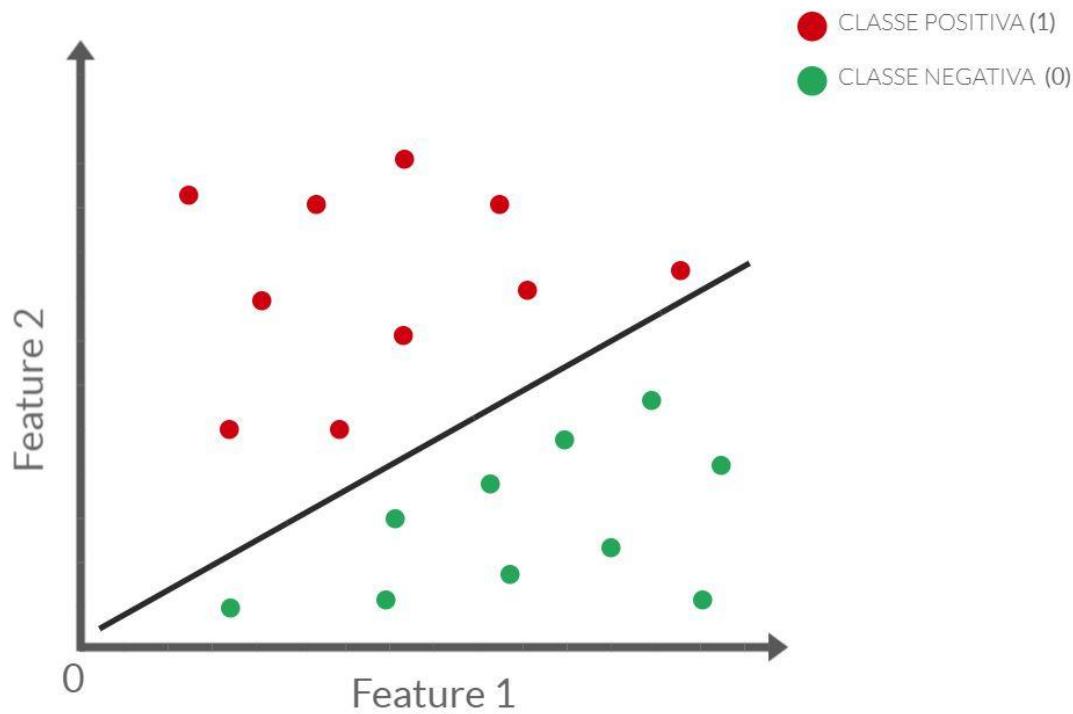
La Classificazione Binaria

presentato da
Giuseppe Gullo

La Classificazione

E' un task dell'apprendimento supervisionato, in cui l'oggetto della previsione è una classe, quindi una variabile categorica.

La Classificazione



Classificare un tumore come maligno

Raggio medio	Punti di concavità	Diagnosi
17.99	0.14710	maligno
13.540	0.047810	benigno
13.080	0.031100	benigno
20.57	0.07017	maligno
9.504	0.020760	benigno
19.69	0.12790	maligno
...
8.196	0.005917	benigno

Classificare un tumore come maligno

Raggio medio	Punti di concavità	Diagnosi
1.097064	2.532475	maligno
-0.166799	-0.028609	benigno
-0.297446	-0.726337	benigno
1.829821	0.548144	maligno
-1.313080	-0.726337	benigno
1.579888	2.037231	maligno
...
1.684571	-1.109197	benigno

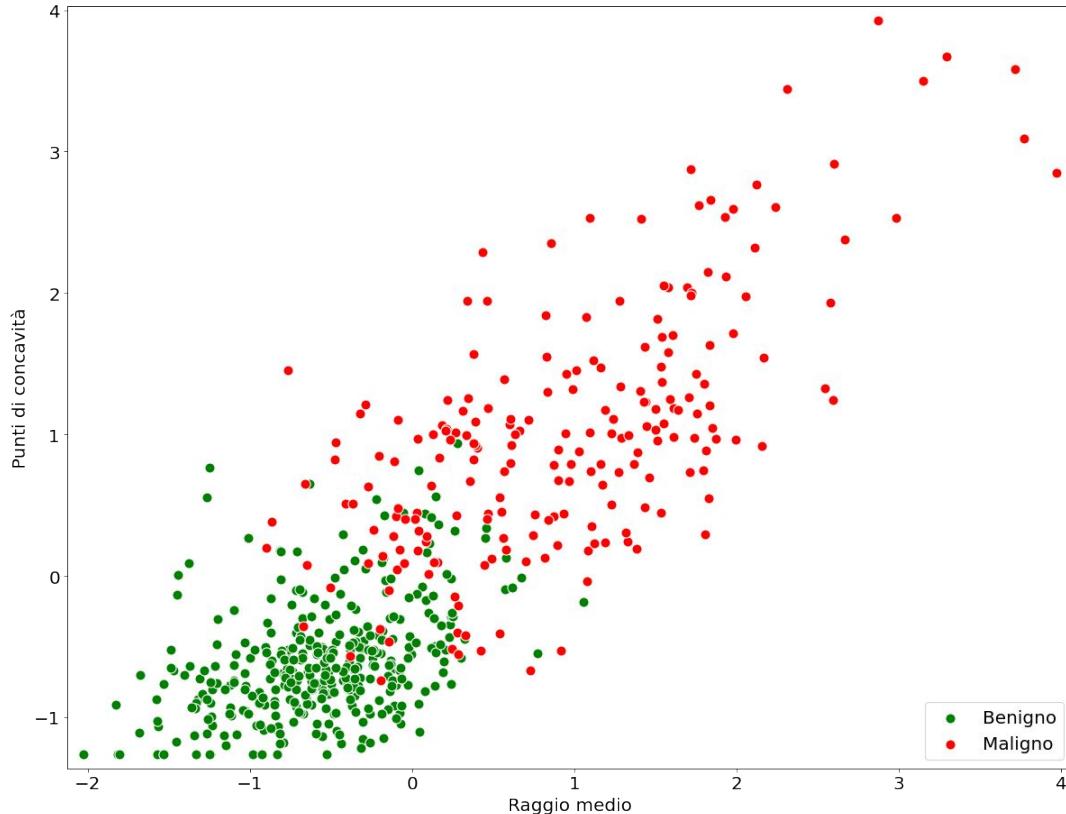
Classificare un tumore come maligno

Raggio medio	Punti di concavità	Diagnosi
1.097064	2.532475	1
-0.166799	-0.028609	0
-0.297446	-0.726337	0
1.829821	0.548144	1
-1.313080	-0.726337	0
1.579888	2.037231	1
...
1.684571	-1.109197	0

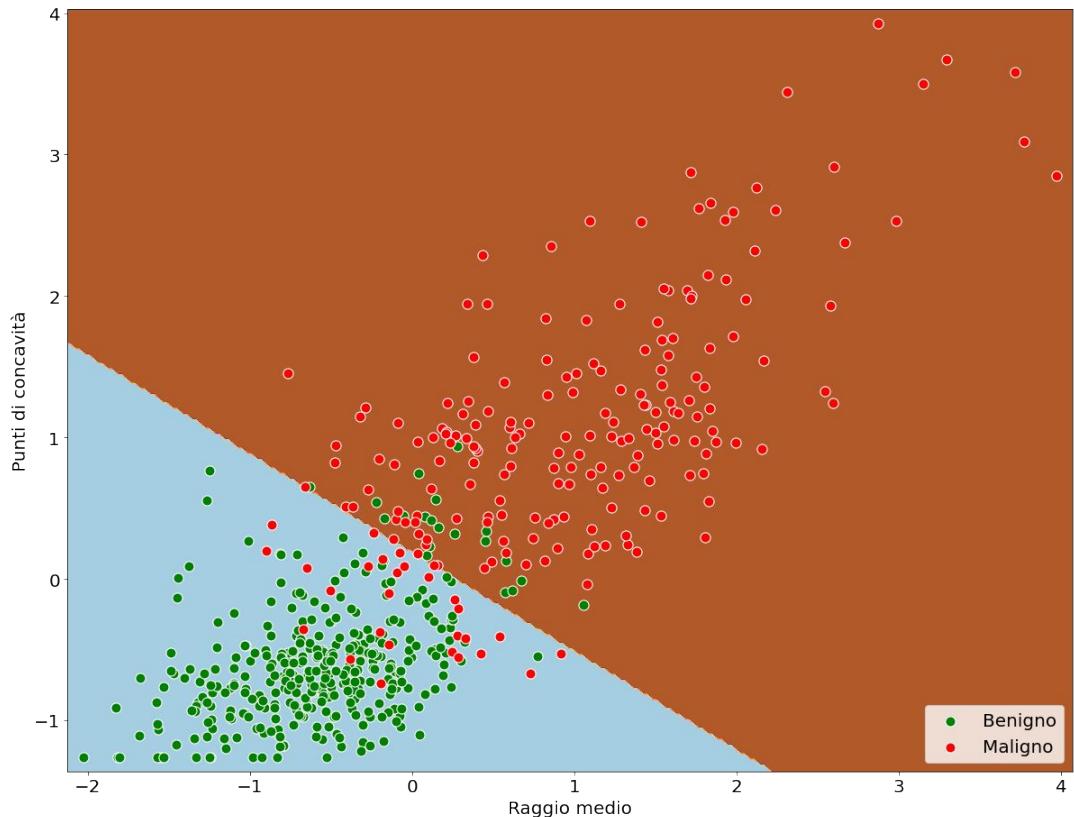
Maligno = Classe Positiva = 1

Benigno = Classe Negativa = 0

Classificare un tumore come maligno



Classificare un tumore come maligno

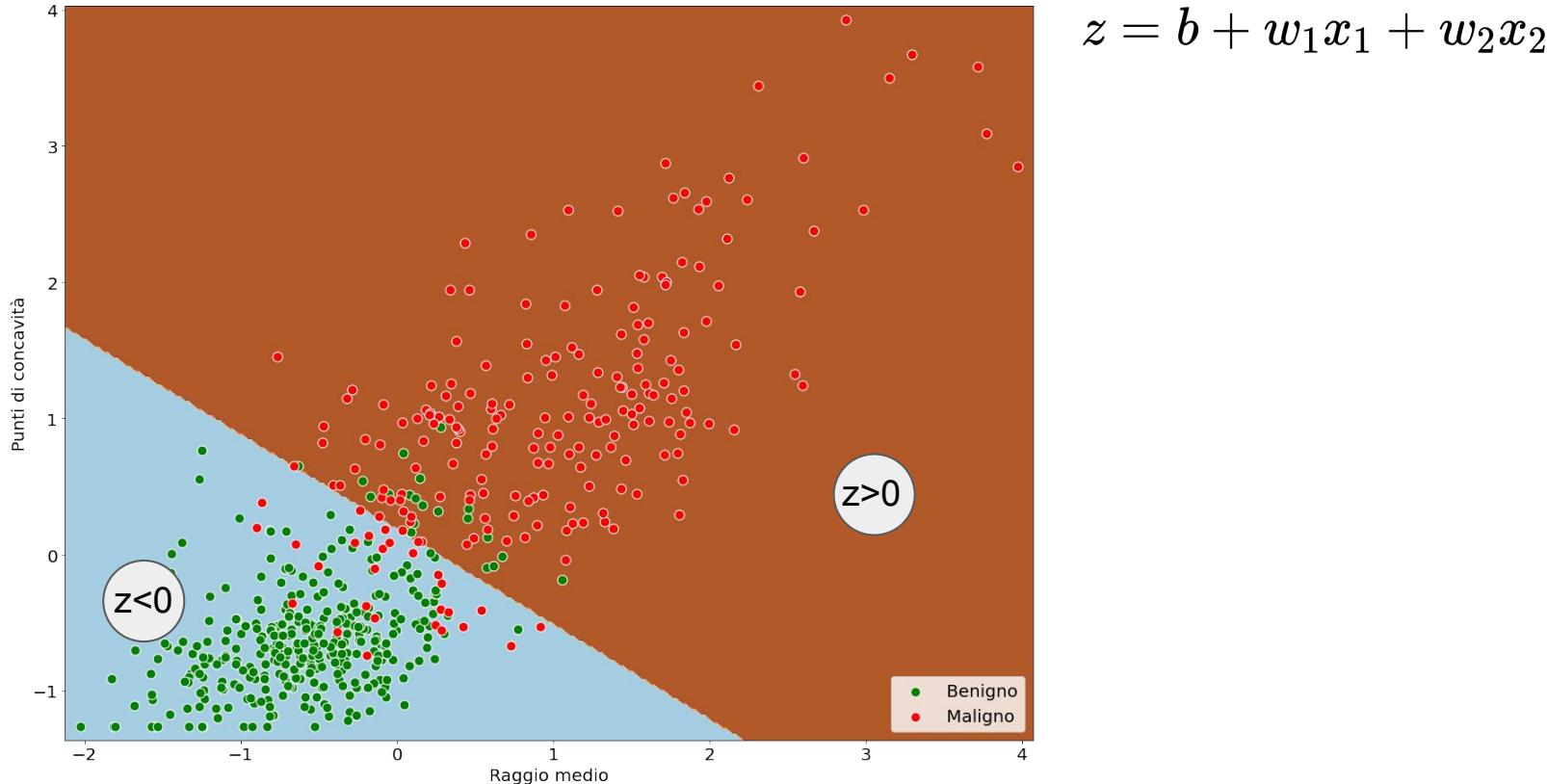


Il Decision Boundary (Soglia di decisione)

E' la retta che separa quanto meglio le due classi.

$$b + w_1x_1 + w_2x_2 = 0$$

Classificare un tumore come maligno



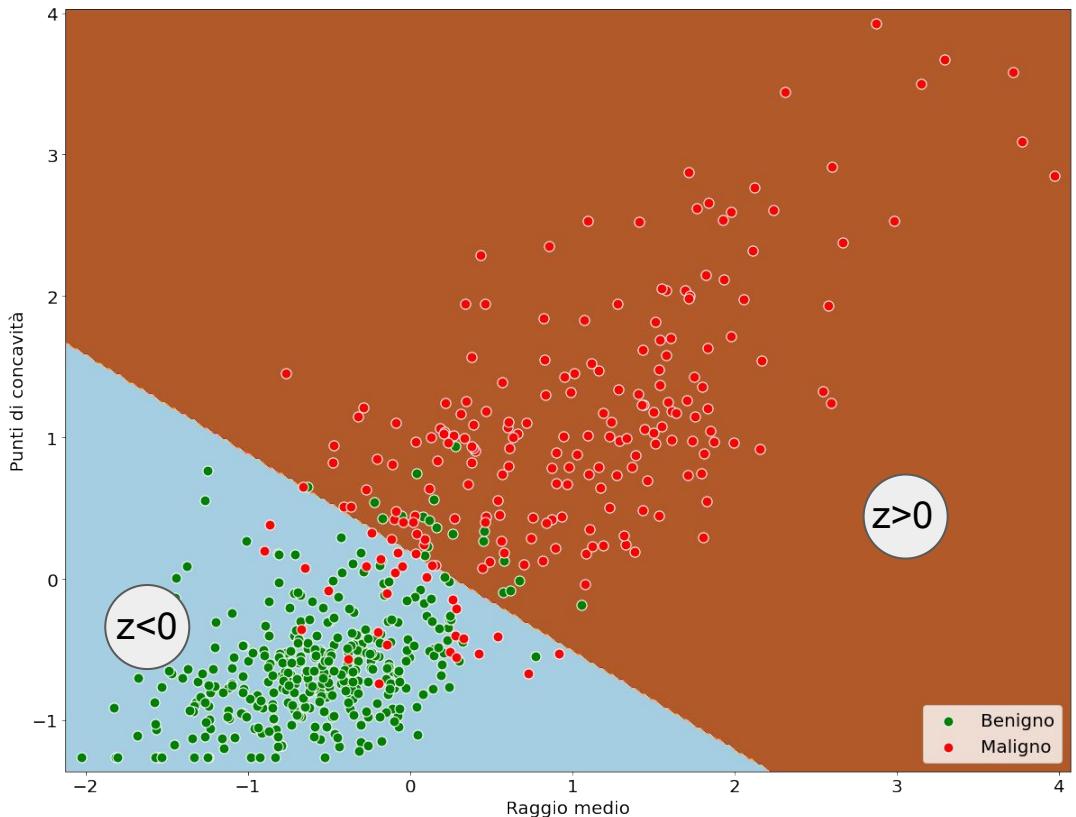
Activation Function (Funzione di Attivazione)

Definisce l'output della classificazione

Binary Step

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases}$$

Classificare un tumore come maligno



$$z = b + w_1 x_1 + w_2 x_2$$

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases}$$

Fondamenti di Machine Learning

La Classificazione

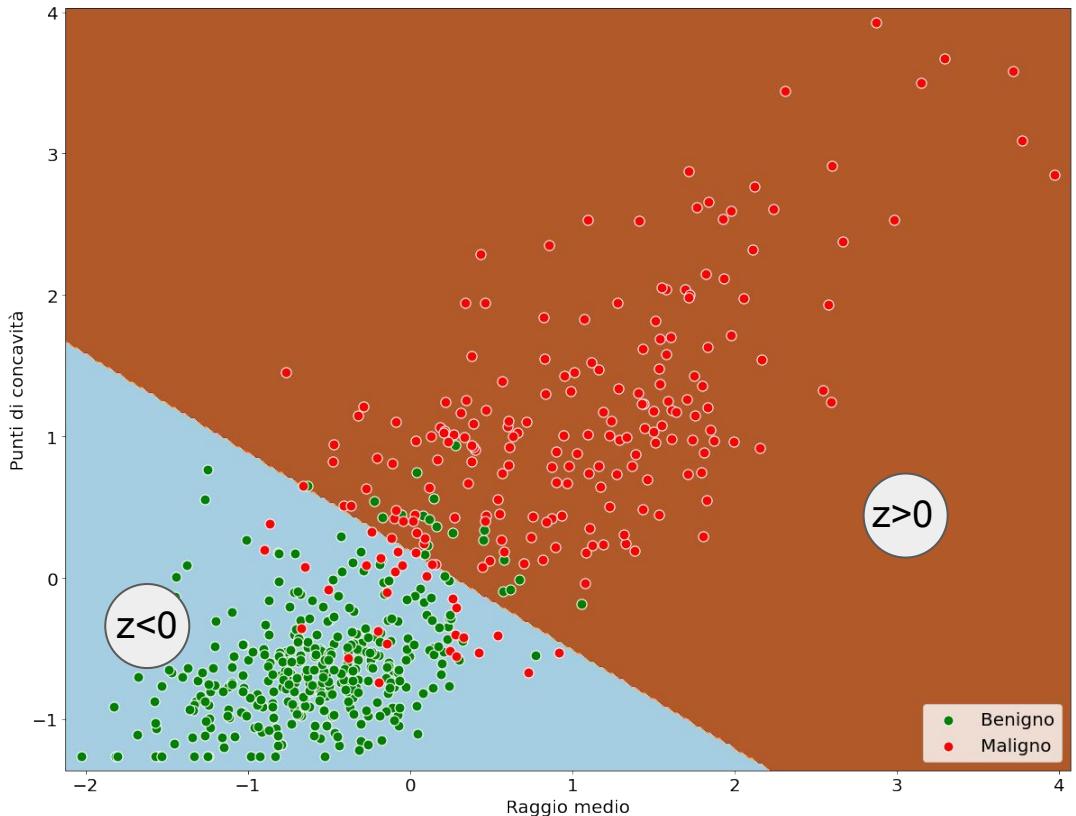
La Regressione Logistica

presentato da
Giuseppe Gullo

La Regressione Logistica

E' un modello di classificazione binaria,
che ci permette di stimare la probabilità
che una data osservazione appartenga alla classe positiva.

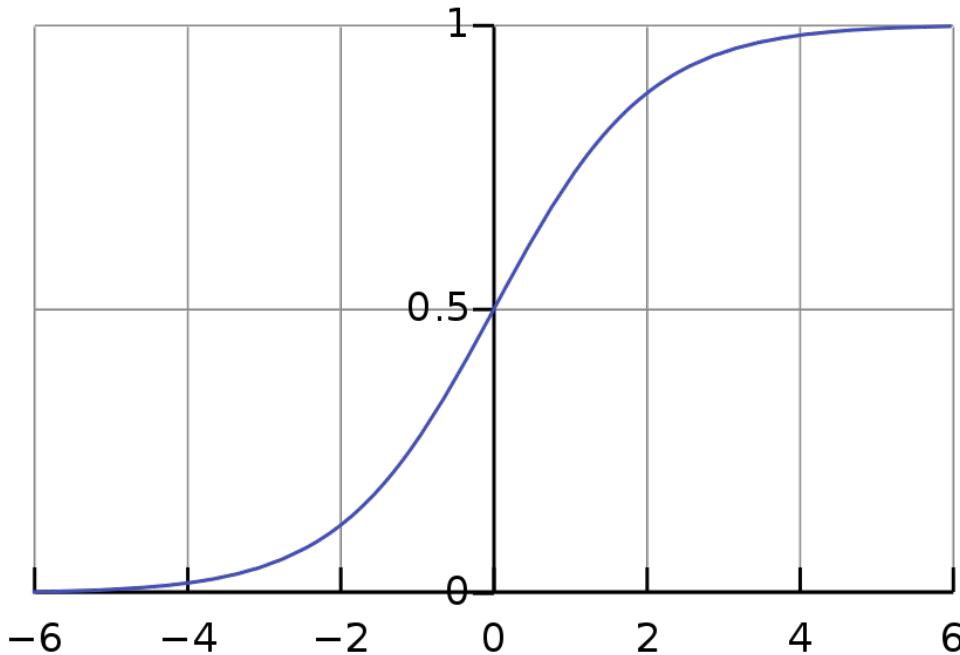
Classificare un tumore come maligno



$$z = b + w_1 x_1 + w_2 x_2$$

$$\phi(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases}$$

Funzione Logistica (a.k.a Sigmoide)



$$\phi(z) = \frac{1}{1 + e^{-z}}$$

Funzione Logistica come Funzione di Attivazione

$$-\infty < z < \infty$$

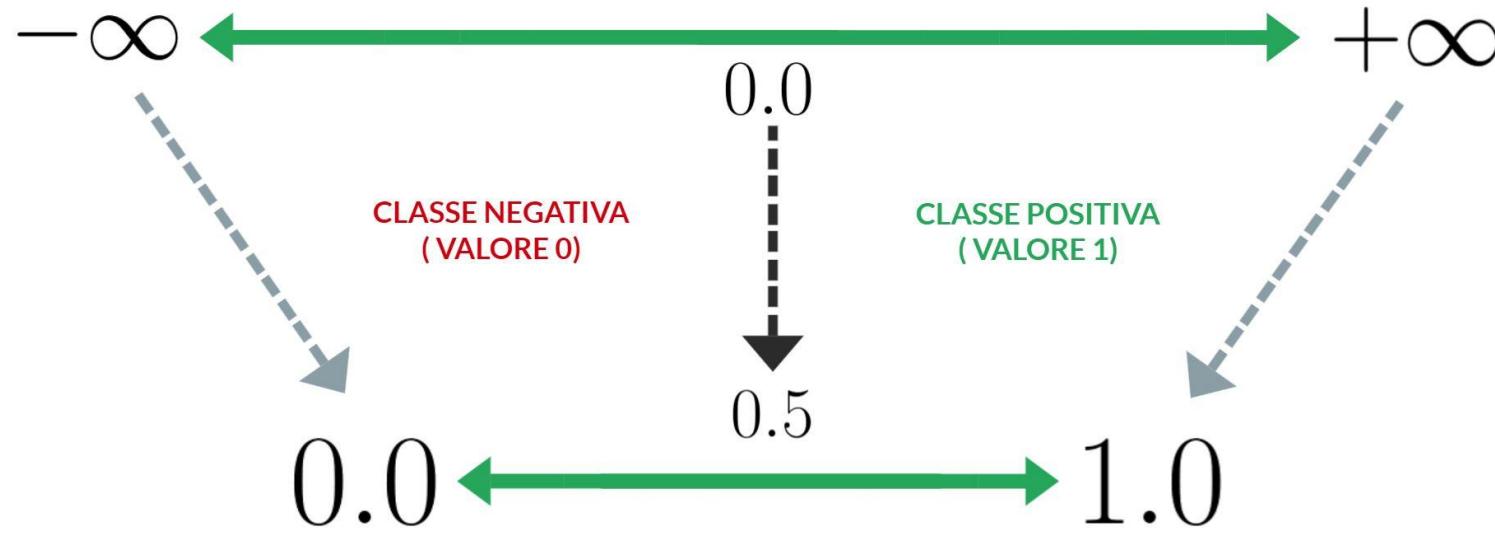


$$\hat{y} = \frac{1}{1 + e^z}$$

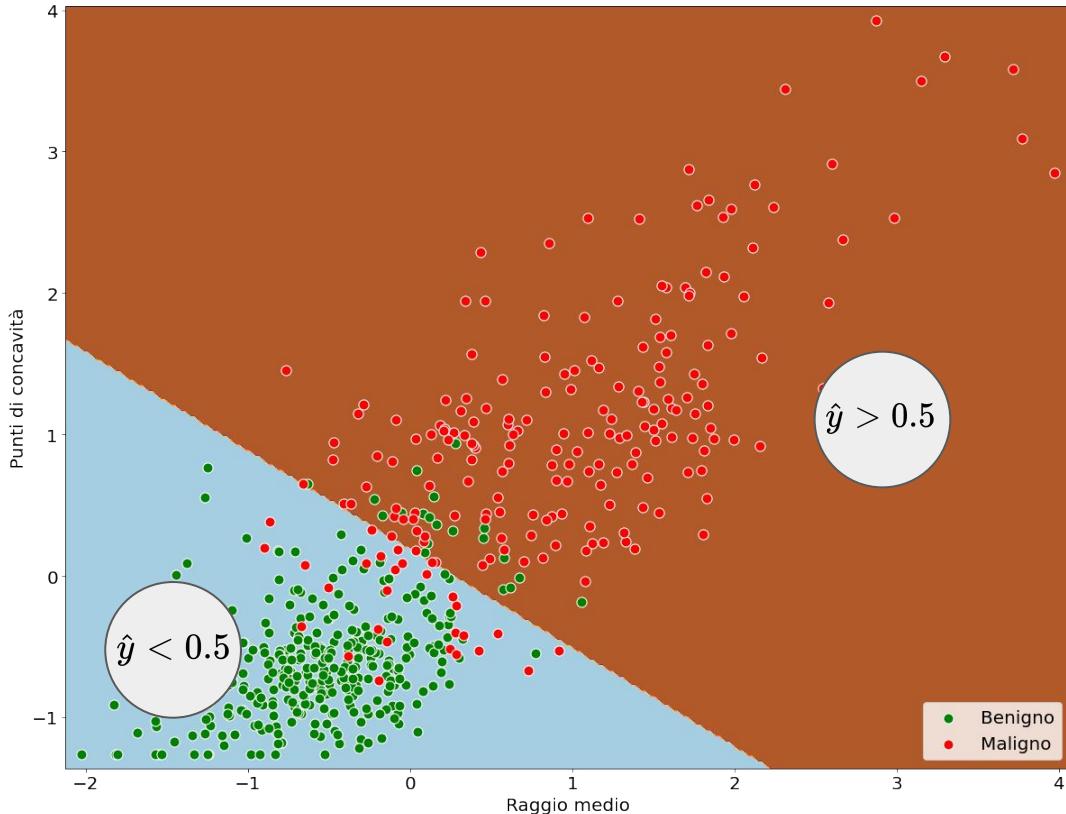


$$0 \leq \hat{y} \leq 1$$

Funzione Logistica come Funzione di Attivazione



Classificare un tumore come maligno



$$z = b + w_1 x_1 + w_2 x_2$$

$$\hat{y} = \frac{1}{1 + e^z}$$

Come si addestra un modello di Regressione Logistica?

Minimizzando la funzione di costo
tramite un algoritmo di ottimizzazione iterativo.

Funzione di costo per la Regressione Logistica

Likelihood

$$L(W) = P(Y|X, W)$$

Qual è la probabilità di ottenere il target Y,
avendo le features X e i parametri W?

Funzione di costo per la Regressione Logistica

Likelihood

$$L(W) = P(Y|X, W)$$

Qual è la probabilità di ottenere il target Y,
avendo le features X e i parametri W?

Funzione di costo per la Regressione Logistica

Log Likelihood

$$l(W) = \log L(W)$$

L'utilizzo del logaritmo rende la fase di addestramento più semplice per la macchina.

Funzione di costo per la Regressione Logistica

Log Loss (a.k.a Cross Entropy)

$$J(W) = -l(W)$$

Possiamo massimizzare la Log Likelihood
o minimizzare la Log Loss.

Fondamenti di Machine Learning

La Classificazione

La Classificazione Multiclasse

presentato da
Giuseppe Gullo

La Classificazione Multiclasse

E' un caso di classificazione in cui le classi oggetto della classificazione sono due o più.

Metodo One vs Rest (OvR)

a.k.a One vs All (OvA)

Ci permette di affrontare un problema di classificazione multiclasse come più problemi di classificazione binaria.

Metodo One vs Rest (OvR)

L'obiettivo è addestrare un modello per ogni classe.



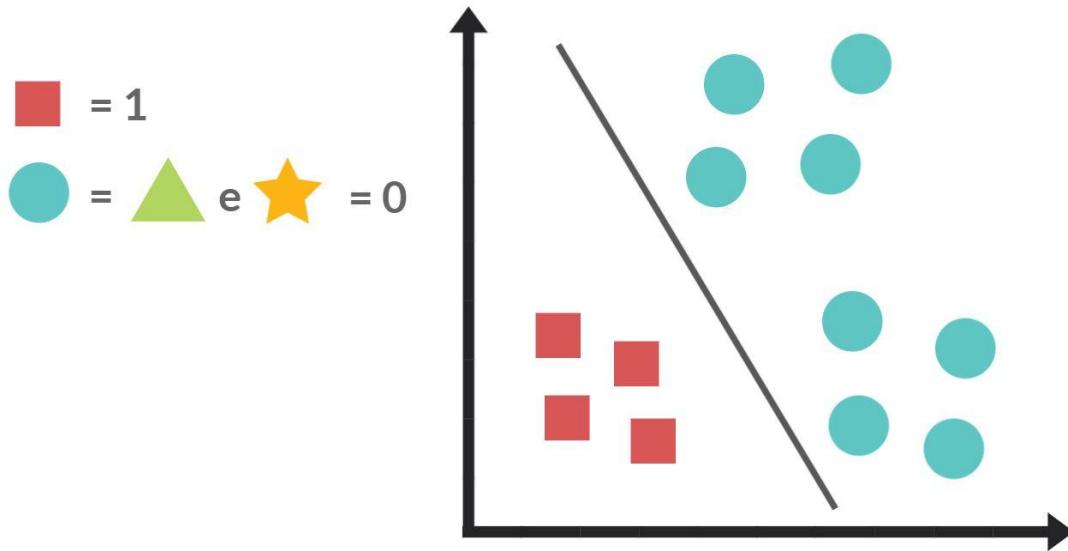
Metodo One vs Rest (OvR)

Per ogni modello, consideriamo una classe come la classe positiva e tutte le altre classi come la classe negativa.



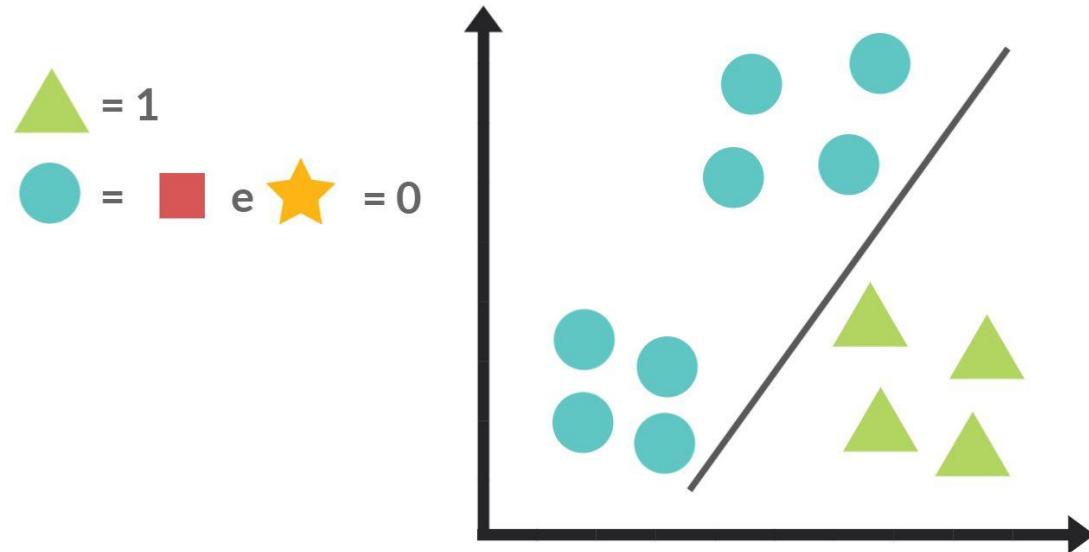
Metodo One vs Rest (OvR)

Addestriamo un modello di regressione logistica
in grado di distinguere le due classi



Metodo One vs Rest (OvR)

Ripetiamo il processo per ogni classe.



Metodo One vs Rest (OvR)

Ripetiamo il processo per ogni classe.



Metodo One vs Rest (OvR)

Al termine avremo ottenuto un modello per ogni classe

$$P(Y = \blacksquare | X, W)$$

$$P(Y = \blacktriangle | X, W)$$

$$P(Y = \star | X, W)$$

Metodo One vs Rest (OvR)

Per eseguire una classificazione, usiamo tutti i modelli.



$$P(Y = \blacksquare | X, W) = 0.5$$

$$P(Y = \blacktriangle | X, W) = 0.3$$

$$P(Y = \star | X, W) = 0.2$$

Metodo One vs Rest (OvR)

Prendiamo per buona la classificazione con la probabilità più alta.

$$P(Y = \blacksquare | X, W) = 0.5$$

$$P(Y = \blacktriangle | X, W) = 0.3$$

$$P(Y = \blacksquare | X, W) = 0.2$$

Metodo One vs One (OvO)

Anche questo ci permette di affrontare un problema di classificazione multiclasse come più problemi di classificazione binaria.

In questo caso addestriamo un modello per ogni classe contro ogni altra classe

Metodo One vs One (OvO)

L'obiettivo è addestrare un modello per ogni classe contro ogni altra classe.

Modello 1  vs 

Modello 2  vs 

Modello 3  vs 

Metodo One vs One (OvO)

L'obiettivo è addestrare un modello per ogni classe contro ogni altra classe.

Modello 1  vs 

Modello 2  vs 

Modello 3  vs 

Prendiamo sempre per buona la classificazione con la probabilità più alta.

Metodo One vs One (OvO)

All'aumentare del numero di classi, il numero di modelli da addestrare aumenta molto più velocemente rispetto all OVR.

Modelli nel OVR = Numero di Classi

Modelli nel OVO = $(\text{Numero di Classi} * (\text{Numero di Classi} - 1)) / 2$

Classi e Label

Un'osservazione può essere assegnata ad una sola classe (mutualmente esclusive)

ES. Se sono nato in Italia non posso essere nato anche negli Stati Uniti

Un'osservazione può essere assegnata a più label

ES. Se ho la cittadinanza Italiana posso avere anche quella Statunitense.

Classi e Label

Un'osservazione può essere assegnata ad una sola classe (mutualmente esclusive)

ES. Se sono nato in Italia non posso essere nato anche negli Stati Uniti

Un'osservazione può essere assegnata a più label

ES. Se ho la cittadinanza Italiana posso avere anche quella Statunitense.

Nel caso in cui la variabile target sia un'insieme di label si parla di classificazione multilabel.

Riconoscere numeri scritti a mano



0



1



2



3



4



5



6



7



8



9

Riconoscere numeri scritti a mano



Riconoscere numeri scritti a mano



Riconoscere numeri scritti a mano



X =

	pixel_1	pixel_2	pixel_3	pixel_4	pixel_5	pixel_6	pixel_7	pixel_8	pixel_9	pixel_10	...	pixel_60	pixel_61	pixel_62	pixel_63	pixel_64
0	0.0	0.0	5.0	13.0	9.0	1.0	0.0	0.0	0.0	0.0	...	13.0	10.0	0.0	0.0	0.0
1	0.0	0.0	0.0	12.0	13.0	5.0	0.0	0.0	0.0	0.0	...	11.0	16.0	10.0	0.0	0.0
2	0.0	0.0	0.0	4.0	15.0	12.0	0.0	0.0	0.0	0.0	...	3.0	11.0	16.0	9.0	0.0
3	0.0	0.0	7.0	15.0	13.0	1.0	0.0	0.0	0.0	8.0	...	13.0	13.0	9.0	0.0	0.0
4	0.0	0.0	0.0	1.0	11.0	0.0	0.0	0.0	0.0	0.0	...	2.0	16.0	4.0	0.0	0.0
5	0.0	0.0	12.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	...	16.0	16.0	10.0	0.0	0.0

Y =

classe
9
9
9
9
9

Riconoscere numeri scritti a mano

Abbiamo 10 classi, da 0 a 9.

0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9

Fondamenti di Machine Learning

La Classificazione

Metriche per la Classificazione

presentato da
Giuseppe Gullo

Metriche per la Classificazione

Ci permettono di quantificare la bontà
di un modello di classificazione.

Matrice di Confusione

Ci permette di sapere non solo quanti errori ha commesso il modello, ma anche quali. Riassume le performance del modello

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Matrice di Confusione

Ci permette di sapere non solo quanti errori ha commesso il modello, ma anche quali. Riassume le performance del modello

- TP (True Positive) = Osservazioni positive classificate come positive.
- FP (False Positive) = Osservazioni negative classificate come positive.
- FN (False Negative) = Osservazioni positive classificate come negative.
- TN (True Negative) = Osservazioni negative classificate come negative.

Accuracy

Rappresenta la percentuale di classificazioni corrette.

$$Accuracy = \frac{\text{Numero di classificazioni corrette}}{\text{Numero di classificazioni totali}}$$

Accuracy

Rappresenta la percentuale di classificazioni corrette.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision (Precisione)

Rappresenta la percentuale delle classificazioni positive che erano effettivamente positive.

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

Recall (Recupero)

a.k.a Sensitivity (Sensibilità)

Rappresenta la percentuale delle osservazioni positive che sono state classificate come positive.

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

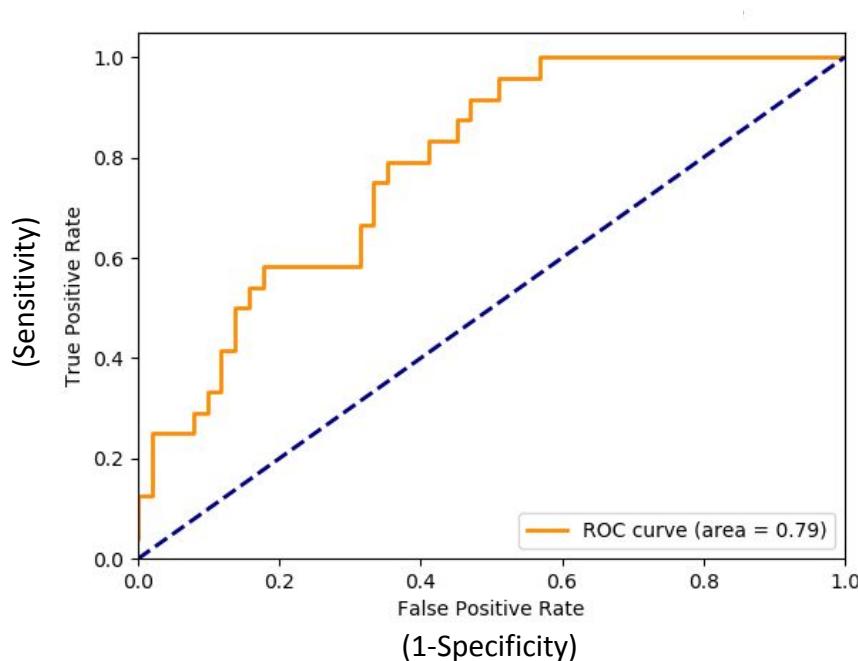
Specificity (Specificità)

Rappresenta la percentuale delle osservazioni negative che sono state classificate come negative.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Receiver Operating Characteristic Curve (ROC)

Ci permette di verificare le performance di un modello di classificazione, per diversi valori di **threshold**.

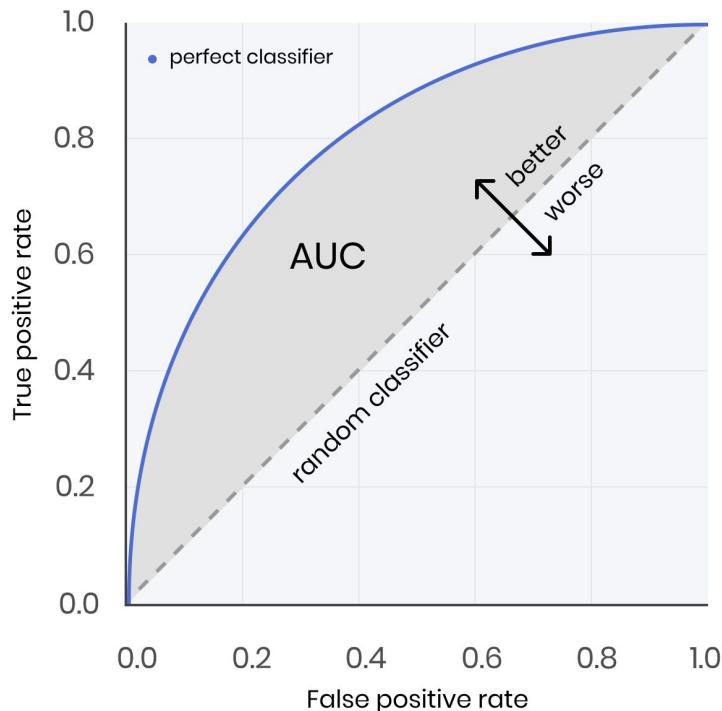


$$\text{True Positive Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

Area Under the ROC Curve (AUC o ROC-AUC)

Ci permette di dare un valore numerico alla curva ROC.



$$\text{True Positive Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

F1 Score

E' la media armonica tra Precision e Recall, sintetizza le due metriche in un unico valore.

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

Log Loss (a.k.a Cross Entropy)

E' una funzione di costo, è la media delle probabilità ritornate dal modello.

$$\text{LogLoss} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$