

Naïve Bayes

Francesco Pugliese, PhD

neural1977@gmail.com

Naive Bayes

- ✓ Naive Bayes è un metodo di classificazione del Machine Learning super efficiente e molto usato recentemente.
- ✓ E' principalmente usato come classificatore di testi nel **Data Mining**.
- ✓ Naive Bayes appartiene ad una famiglia di algoritmi che includono algoritmi sia supervisionati che non supervisionati.
- ✓ I Classificatori Naive Bayes sono una collezione di algoritmi basati sul teorema di Bayes.
- ✓ Non si tratta di un singolo algoritmo ma di una famiglia di algoritmi dove tutti condividono un principio comune, ossia ogni coppia di feature che deve essere classificata è indipendente da ciascuna altra.
- ✓ Ai fini di comprendere Naive Bayes, richiamiamo la regola di Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Il Teorema di Bayes

- ✓ Nel Machine Learning, il nostro scopo è principalmente selezionare la miglior ipotesi ***H*** dati i dati ***d***.
- ✓ In un problema di classificazione, la nostra ipotesi (***H***) potrebbe essere la classe da assegnare ad una nuova istanza di dati (***d***).
- ✓ Uno dei modi più facili di selezionare l'ipotesi più probabile dati i dati che abbiamo è quello di usare la nostra conoscenza a priori circa il problema in questione.
- ✓ **Il Teorema di Bayes** fornisce un modo per calcolare la probabilità dell'ipotesi data la nostra conoscenza a priori:

$$P(H|d) = \frac{P(d|H) * P(H)}{P(d)}$$

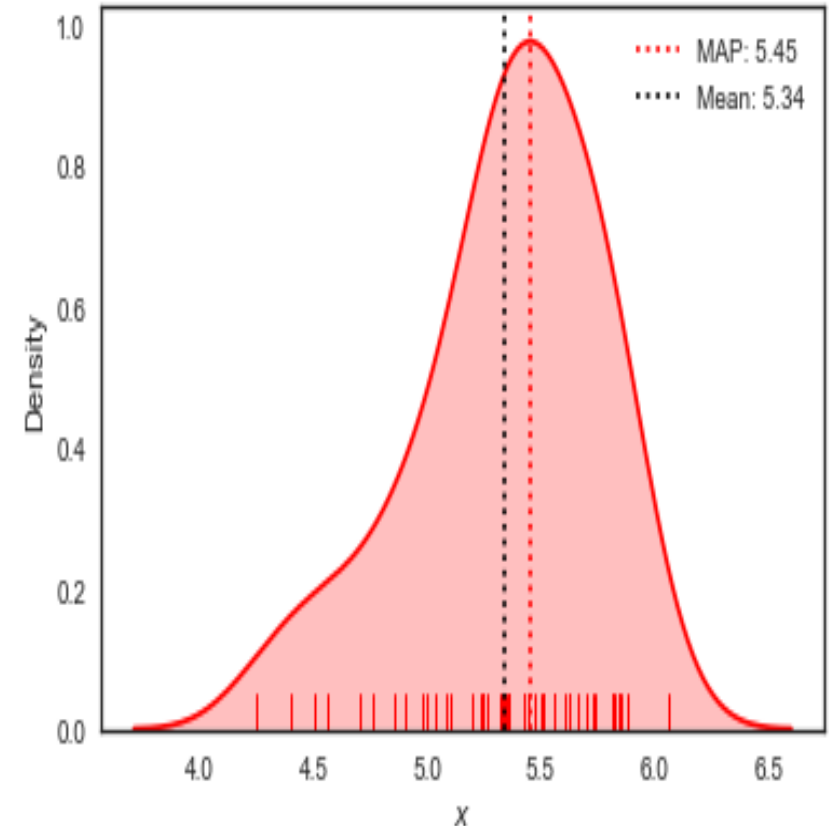
- ✓ Dove: $P(H|d)$ è la "**Probabilità a Posteriori**" dell'ipotesi ***H*** dati i dati ***d***. $P(d|H)$ è la "**Probabilità a Priori**" dei dati ***d*** dato che l'ipotesi ***H*** sia considerata Vera. $P(H)$ è la "**Probabilità a Priori**" che l'ipotesi ***H*** sia vera senza tenere nessuna considerazione dei dati. $P(d)$ è la probabilità dei dati senza alcuna considerazione dell'ipotesi.

Stimatore Maximum a Posteriori (MAP)

- ✓ Dopo aver calcolato la probabilità per un certo numero differente di ipotesi, **Naïve Bayes** seleziona l'ipotesi con la più alta probabilità. Questo è la **Maximum Likely Hypothesis** e potrebbe formalmente essere chiamata la Maximum a Posteriori (**MAP**) hypothesis o **MAP estimator**.

$$MAP(H) = \max(P(H|d)) = \max\left(\frac{P(d|H)*P(H)}{P(d)}\right) = \max(P(d|H) * P(H))$$

- ✓ Possiamo rimuovere **$P(d)$** dal momento che abbiamo l'ipotesi più probabile in assoluto dal momento che questa probabilità è costante ed è usata solo per normalizzare.



Come lavora Naive Bayes? Un esempio

- ✓ Nel Naive Bayes (NB), per ciascun element all'interno del **test set**, il Teorema di Bayes ci permette di determinare la probabilità che l'element appartenga ad una classe.
- ✓ Consideriamo la “**decisione**” (classificatione) di giocare a golf o no (classe binaria si / no) a seconda delle caratteristiche ambientali (features):

	Outlook	Temp	Humidity	Windy	Play
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes
10	Sunny	Mild	Normal	False	Yes
11	Rainy	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Sunny	Mild	High	True	No

Come lavora il Naive Bayes? Un esempio

- ✓ Le Feature sono 4: **Outlook**, **Temp**, **Humidity**, **Windy**. Ora costruiamo la **Tabella** Delle **Frequenze**:

		Play		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
		9	5	

		Play		
		Yes	No	
Temp	Cool	3	1	4
	Mild	4	2	6
	Hot	2	2	4
		9	5	

		Play		
		Yes	No	
Humidity	Normal	6	1	7
	High	3	4	7
		9	5	

		Play		
		Yes	No	
Windy	False	6	2	8
	True	3	3	6
		9	5	

Come lavora il Naive Bayes? Un esempio

✓ Poi, trasformiamo le frequenze in probabilità:

		Play		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

		Play		
		Yes	No	
Temp	Cool	3/9	1/5	4/14
	Mild	4/9	2/5	6/14
	Hot	2/9	2/5	4/14
		9/14	5/14	

		Play		
		Yes	No	
Humidity	Normal	6/9	1/5	7/14
	High	3/9	4/5	7/14
		9/14	5/14	

		Play		
		Yes	No	
Windy	False	6/9	2/5	8/14
	True	3/9	3/5	6/14
		9/14	5/14	

Come lavora il Naive Bayes? Un esempio

- ✓ In questo modo, abbiamo tutte le informazioni adatte a calcolare **probabilità semplici e probabilità condizionali**.
- ✓ Questo significa che il modello è «**Addestrato**» il che significa che il modello è ormai creato, allo stesso modo di ciò che avviene per gli Alberi di Decisione e per il Random Forest.

The diagram illustrates the Naive Bayes model by showing four tables, each representing a different feature (Outlook, Temp, Humidity, Windy) and its relationship to the 'Play' variable (Yes/No). Each table includes counts and probabilities for each outcome.

Outlook Table:

		Play		
		Yes	No	
Outlook	Sunny	3 / 9	2 / 5	5 / 14
	Overcast	4 / 9	0 / 5	4 / 14
	Rainy	2 / 9	3 / 5	5 / 14
		9 / 14	5 / 14	

Probabilities: $p(\text{Sunny}|\text{Yes})$, $p(\text{Overcast}|\text{Yes})$, $p(\text{Rainy}|\text{Yes})$, $p(\text{Sunny}|\text{No})$, $p(\text{Overcast}|\text{No})$, $p(\text{Rainy}|\text{No})$, $p(\text{Yes})$, $p(\text{No})$, $p(\text{Sunny})$, $p(\text{Overcast})$, $p(\text{Rainy})$.

Temp Table:

		Play		
		Yes	No	
Temp	Cool	3 / 9	1 / 5	4 / 14
	Mild	4 / 9	2 / 5	6 / 14
	Hot	2 / 9	2 / 5	4 / 14
		9 / 14	5 / 14	

Probabilities: $p(\text{Cool}|\text{Yes})$, $p(\text{Mild}|\text{Yes})$, $p(\text{Hot}|\text{Yes})$, $p(\text{Cool}|\text{No})$, $p(\text{Mild}|\text{No})$, $p(\text{Hot}|\text{No})$, $p(\text{Yes})$, $p(\text{No})$, $p(\text{Cool})$, $p(\text{Mild})$, $p(\text{Hot})$.

Humidity Table:

		Play		
		Yes	No	
Humidity	Normal	6 / 9	1 / 5	7 / 14
	High	3 / 9	4 / 5	7 / 14
		9 / 14	5 / 14	

Probabilities: $p(\text{Normal}|\text{Yes})$, $p(\text{High}|\text{Yes})$, $p(\text{Normal}|\text{No})$, $p(\text{High}|\text{No})$, $p(\text{Yes})$, $p(\text{No})$, $p(\text{Normal})$, $p(\text{High})$.

Windy Table:

		Play		
		Yes	No	
Windy	False	6 / 9	2 / 5	8 / 14
	True	3 / 9	3 / 5	6 / 14
		9 / 14	5 / 14	

Probabilities: $p(\text{True}|\text{Yes})$, $p(\text{False}|\text{Yes})$, $p(\text{True}|\text{No})$, $p(\text{False}|\text{No})$, $p(\text{Yes})$, $p(\text{No})$, $p(\text{True})$, $p(\text{False})$.

Come lavora il Naive Bayes? Un esempio

- ✓ Consideriamo l'esempio di test: **Outlook**=Rainy, **Temp**=Cool, **Humidity**=High, **Windy**=True.
- ✓ L'algoritmo calcola la decisione di giocare determinando le probabilità condizionali a posteriori e scegliendo quella massima ossia lo stimatore **MAP**:

$$P(Yes | X) = P(Rainy | Yes) \cdot P(Cool | Yes) \cdot P(High | Yes) \cdot P(True | Yes) \cdot P(Yes)$$
$$P(No | X) = P(Rainy | No) \cdot P(Cool | No) \cdot P(High | No) \cdot P(True | No) \cdot P(No)$$

- ✓ Sostituiamo i simboli con i valori:

$$P(Yes | X) = (2/9) \cdot (3/9) \cdot (3/9) \cdot (3/9) \cdot (9/14)$$
$$P(No | X) = (3/5) \cdot (1/5) \cdot (4/5) \cdot (3/5) \cdot (5/14)$$

Come lavora il Naive Bayes? Un esempio

✓ Che sono:

$$P(Y e s | X) = 0.00529$$

$$P(N o | X) = 0.02057$$

✓ Normalizzando:

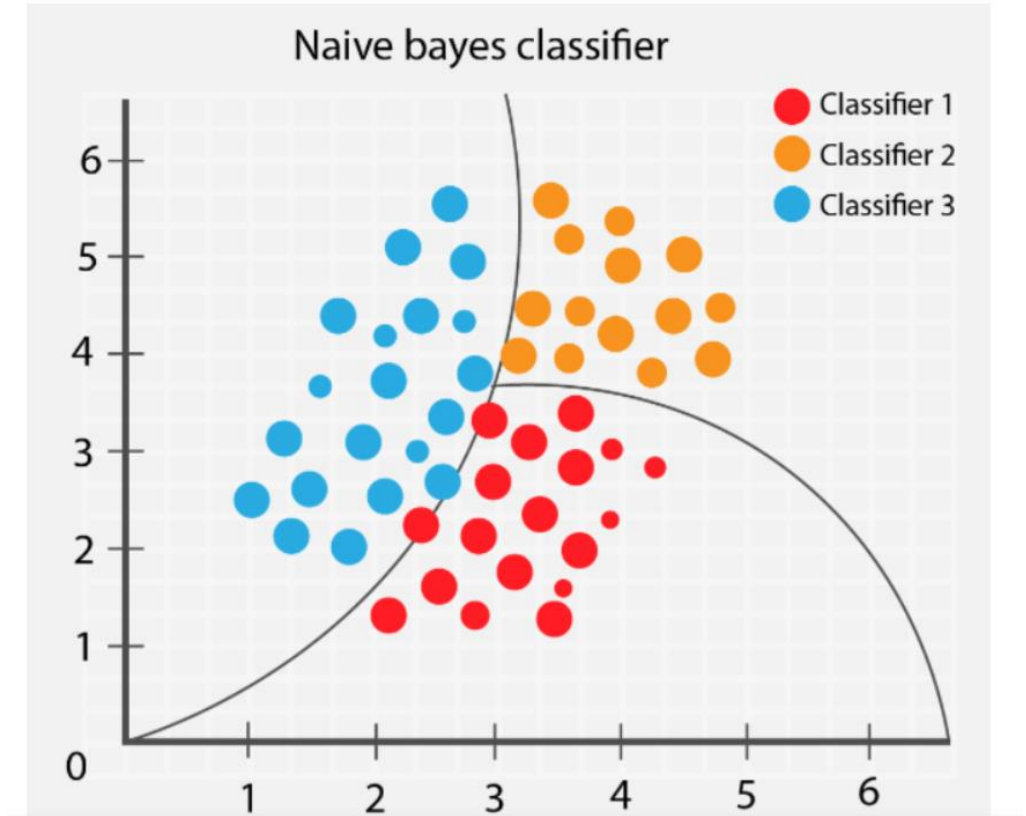
$$P(Y e s | X) = 0.00529 / 0.00529 + 0.02057$$

$$P(N o | X) = 0.02057 / 0.00529 + 0.02057$$

✓ E quindi:

$$P(Y e s | X) = 0.20$$

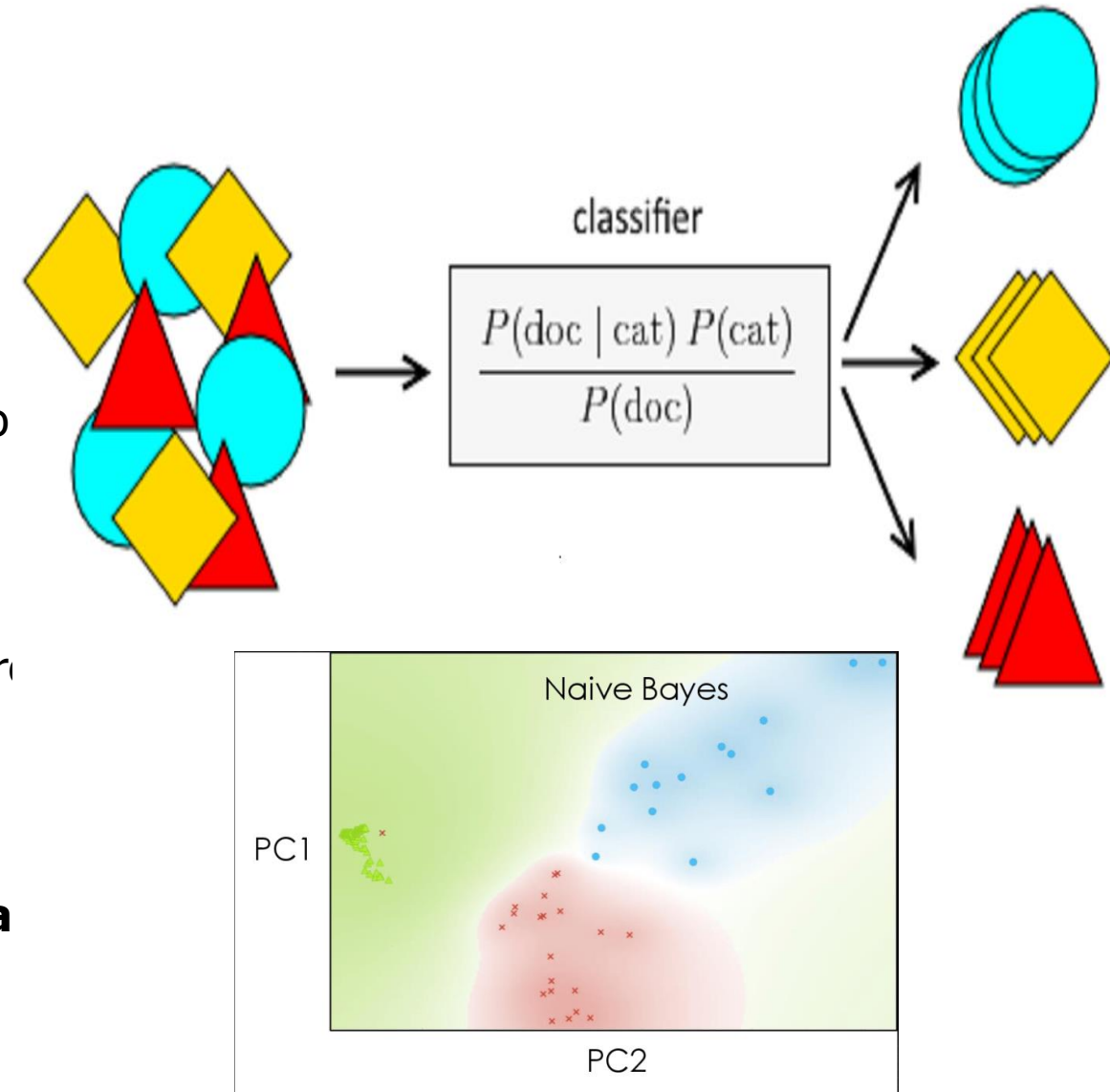
$$P(N o | X) = 0.80$$



✓ Che significa: a seconda di queste feature fornite, la probabilità di giocare a golf è **20%**, mentre quella di non giocare è **80%**. **No** è la risposta del classificatore NB dunque.

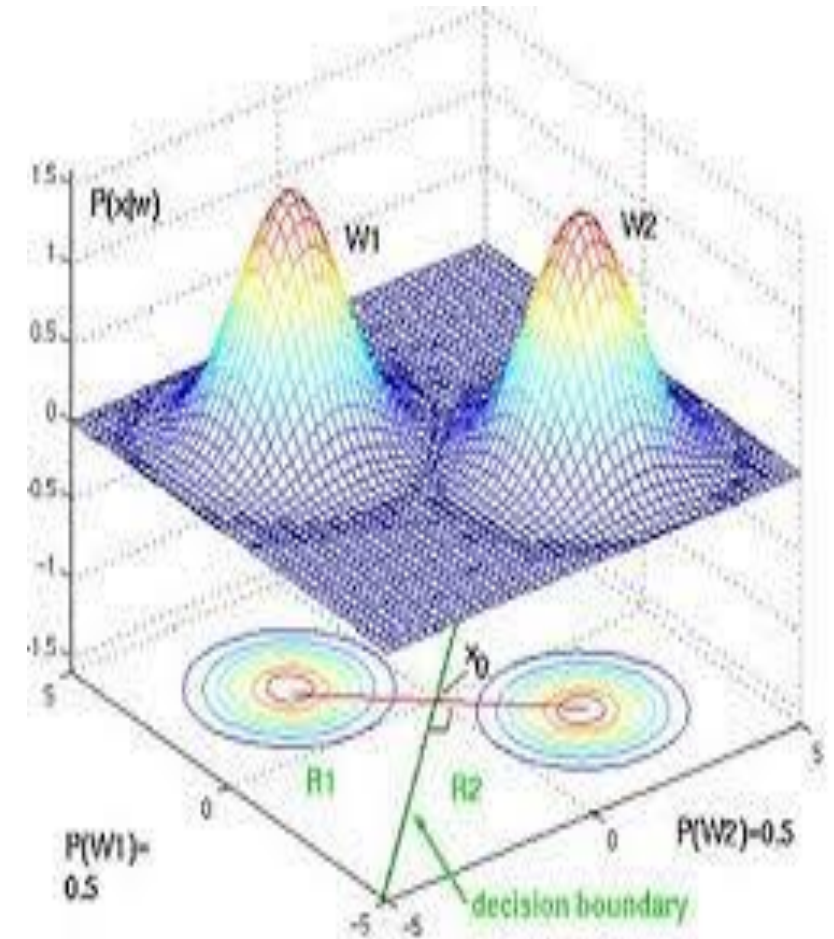
Che cosa è "naive" (banale, semplice) : in Naive Bayes?

- ✓ Naive Bayes (NB) è **ingenuo** perchè fa **l'assunzione** che gli attributi della misurazione siano **indipendenti** tutti da ciascun altro. .
- ✓ Possiamo semplicemente **considerare** un attributo come quantità indipendente e determinare la proporzione delle precedenti misurazioni che appartengono a quella classe che ha lo stesso valore per questo attributo solamente.
- ✓ Naive Bayes è usato principalmente **per predire la probabilità di differenti classi basate su attribute multipli.**



Gaussian Naive Bayes

- ✓ Prima, abbiamo calcolato le probabilità dei valori di input per ciascuna classe utilizzando la frequenza. Con gli input di valori reali, possiamo calcolare la **media** e la **deviazione standard** dei valori di input (x) per ciascuna classe per riassumere la distribuzione.
- ✓ Questo significa che in aggiunta alle probabilità per ciascuna classe, dobbiamo anche immagazzinare la media e la deviazione standard per ciascuna variabile di input di ogni classe.
- ✓ Le probabilità di nuovi valori x vengono calcolati usando la **Funzione Densità di Probabilità Gaussiana (Gaussian PDF)**.
- ✓ Mentre si eseguono le predizioni, questi parametri possono essere inseriti nella **Gaussian PDF** con un nuovo con un nuovo input per la variabile, e di ritorno la Gaussian PDF fornirà una stima della probabilità di quel nuovo valore di input per quella classe.



Applicazioni di Naive Bayes

- ✓ Classificare un nuovo articolo che parli di tecnologia, politica o sport
- ✓ Sentiment analysis sui social media
- ✓ Software di Riconoscimento Facciale
- ✓ Sistemi di Raccomandazione come in Netflix, Amazon
- ✓ Filtraggio dello Spam

Francesco Pugliese

