Naïve Bayes

Francesco Pugliese, PhD

neural1977@gmail.com

We will discuss about…

✓ Naive Bayes

Naive Bayes

- ✓ Naive Bayes is super effective, commonly-used machine learning classifier.
- ✓ It is mostly used in **text classification** while mining the data.
- ✓ Naive Bayes is in its own a family of algorithms including algorithms for both supervised and unsupervised learning.
- ✓ Naive Bayes classifiers are a collection of classification algorithms based on Bayes theorem.
- ✓ It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.
- ✓ In order to understand Naive Bayes, let us recall Bayes rule:

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

Bayes' Theorem

- ✓ In machine learning, our purpose is mainly selecting the best hypothesis ***H*** given data ***d***.
- ✓ In a classification problem, our hypothesis (**H**) may be the class to assign for a new data instance (**d**).
- ✓ One of the easiest ways of selecting the most probable hypothesis given the data that we have that we can use as our **prior knowledge** about the problem.
- ✓ **Bayes' Theorem** provides a way that we can calculate the probability of a hypothesis given our prior knowledge.

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

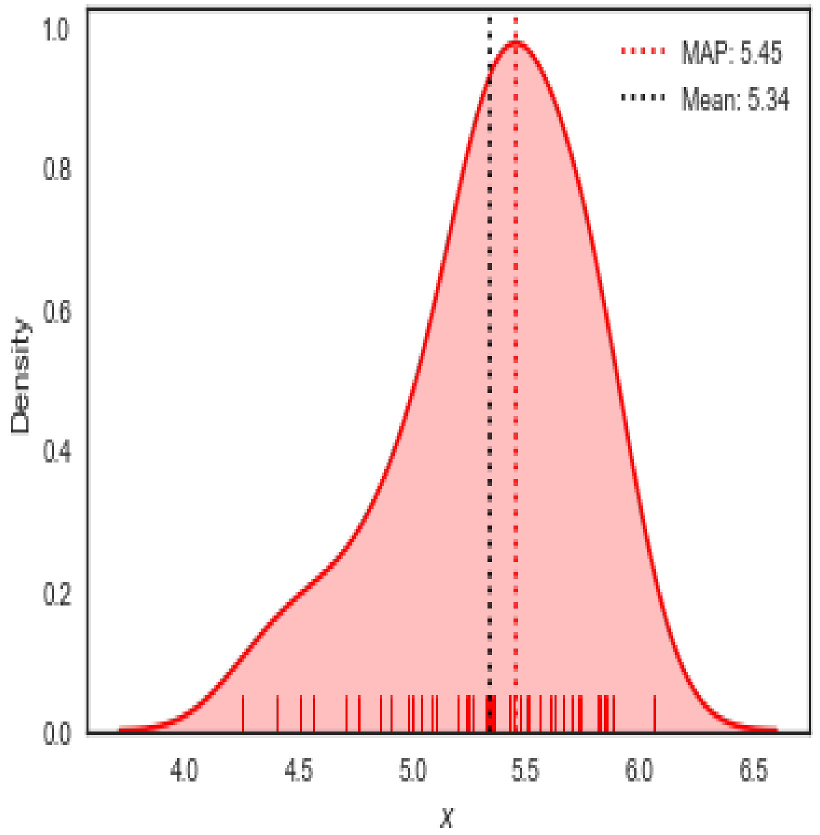
- ✓ Where: $P(H|d)$ is the "**Posterior Probability"** of hypothesis ***H*** given the data ***d***. $P(d|H)$ is the **"Prior Probability"** of the data ***d*** given that the hypothesis ***H*** was true. $P(H)$ is the **"Prior Probability"** of hypothesis ***H*** being true with no consideration of the data. $P(d)$ is the probability of the data with no consideration of the hypothesis).

Maximum a Posteriori Estimator (MAP)

- ✓ After calculating the posterior probability for a number of different hypotheses, Naïve Bayes selects the hypothesis with the highest probability. This is the **Maximum Likely Hypothesis** and may formally be called the Maximum a Posteriori (**MAP**) hypothesis or **MAP estimator**.

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

- ✓ We can remove ***P(d)*** since we need the most probable hypothesis as it is constant and only used to normalize.



How Naive Bayes Works? An Example

- ✓ In Naive Bayes (NB), for each element within the test set, the Bayes' Theorem allows us to determine the probability that element belongs to one class.

- ✓ Let's consider the "**decision**" (classification) of playing golf or not (binary class yes/no) according to environmental characteristics (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

How Naive Bayes Works? An Example

✓ Features are 4: **Outlook, Temp, Humidity, Windy**. We now build the Frequencies Table:

		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	

How Naive Bayes Works? An Example

✓ Then, we transform the frequencies into probabilities:

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

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

How Naive Bayes Works? An Example

- ✓ This way, we have all the information suitable to calculate **simple probabilities** and **conditional probabilities**.

- ✓ This means that the model is **«Trained»** meaning that the model is created same as we do with Decision Trees and Random Forest.

p(Sunny|Yes) p(Overcast|Yes) p(Sunny|No) p(Overcast|No)

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

p(Rainy|Yes) p(Rainy|No)

p(Cool|Yes) p(Cool|No)

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

p(Mild|Yes) p(Hot|Yes) p(Mild|No) p(Hot|No)

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

p(Normal|Yes) p(High|Yes) p(Normal|No) p(High|No)

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

p(True|Yes) p(True|Yes) p(False|No) p(False|No)

How Naive Bayes Works? An Example

- ✓ Let’s consider the test sample: **Outlook**=Rainy, **Temp**=Cool, **Humidity**=High, **Windy**=True.

- ✓ The algorithm calculates the decision to play by determining the following conditional probabilities:

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

- ✓ Let’s subsititute symbols with values:

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

How Naive Bayes Works? An Example

✓ Which is:

$$P(Yes \mid X) = 0.00529$$

$$P(No \mid X) = 0.02057$$

✓ By **normalizing**:

$$P(Yes \mid X) = 0.00529 / 0.00529 + 0.02057$$

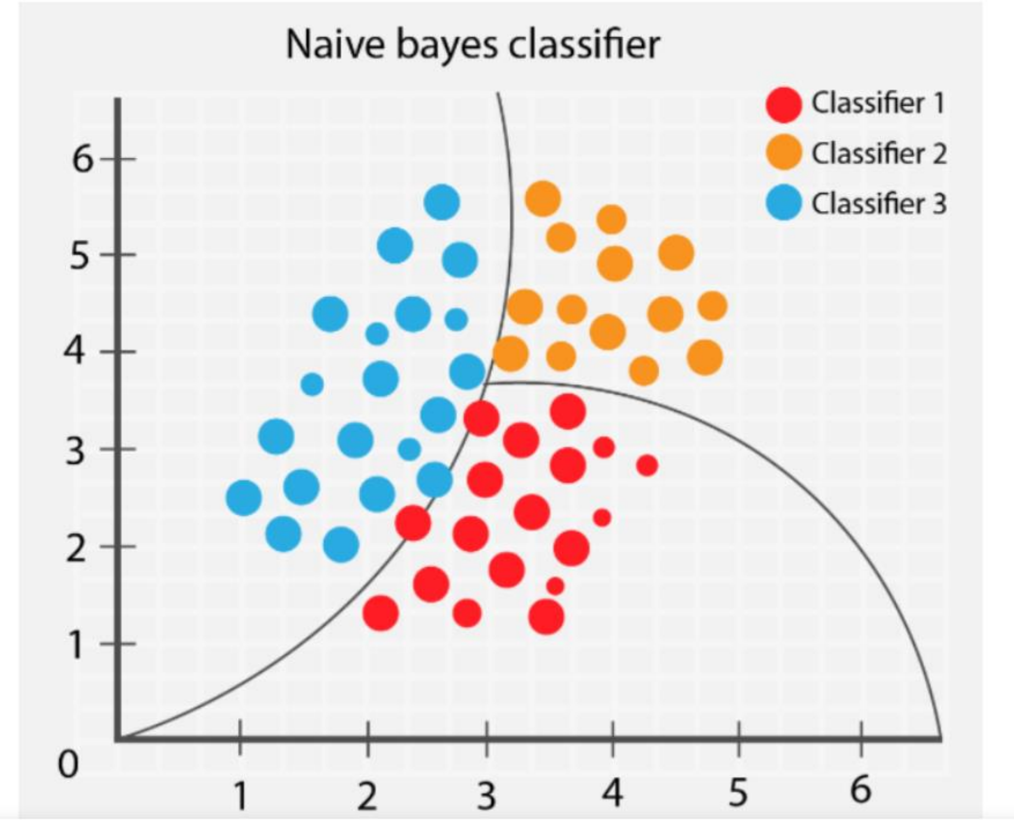
$$P(No \mid X) = 0.02057 / 0.00529 + 0.02057$$

✓And so:

$$P(Yes \mid X) = 0.20$$

$$P(No \mid X) = 0.80$$

✓Which means: based on the features, the likelihood to play golf is **20%,** while **80%** to not play. **No** is the answer of the NB classifier then

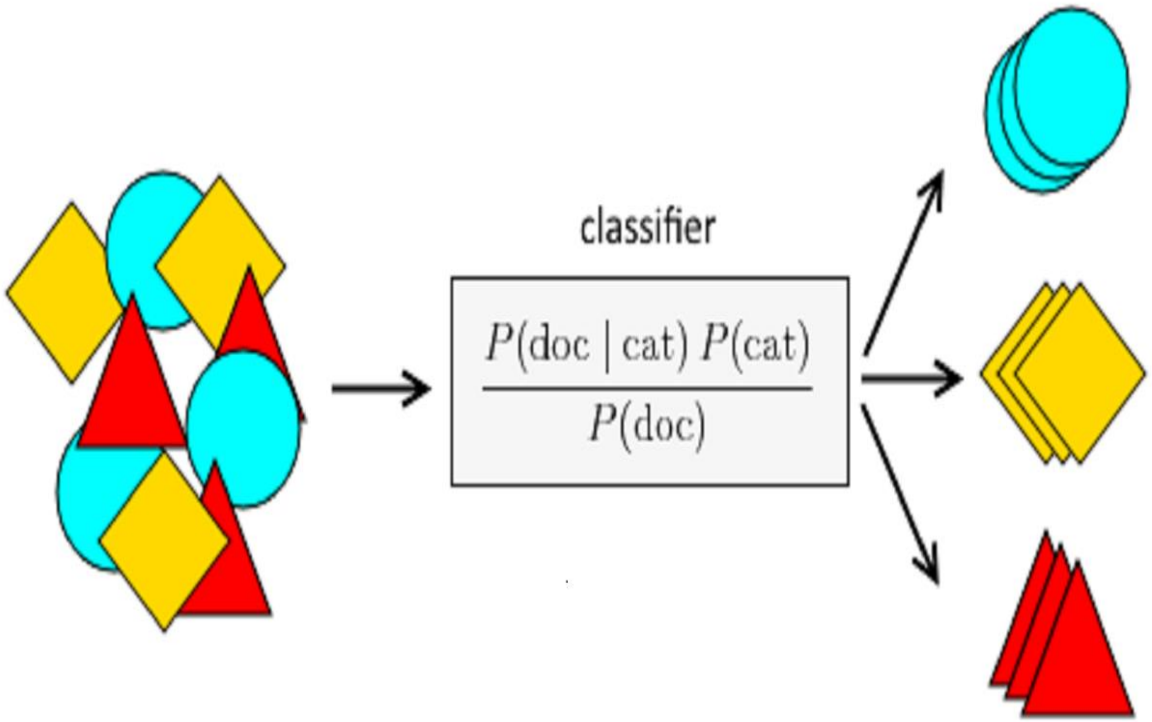


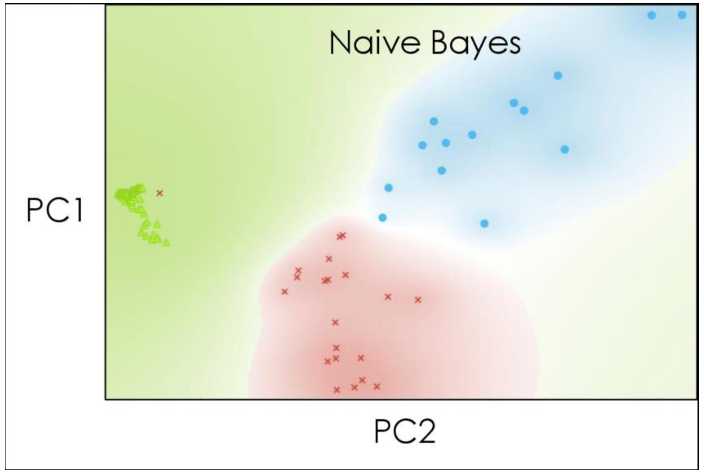
What is so "naive": in Naive Bayes?

- ✓ Naive Bayes (NB) is naive because it makes the **assumption** that attributes of a measurement are **independent** of each other.

- ✓ We can simply **take** one attribute as independent quantity and determine proportion of previous measurements that belong to that class having the same value for this attribute only.

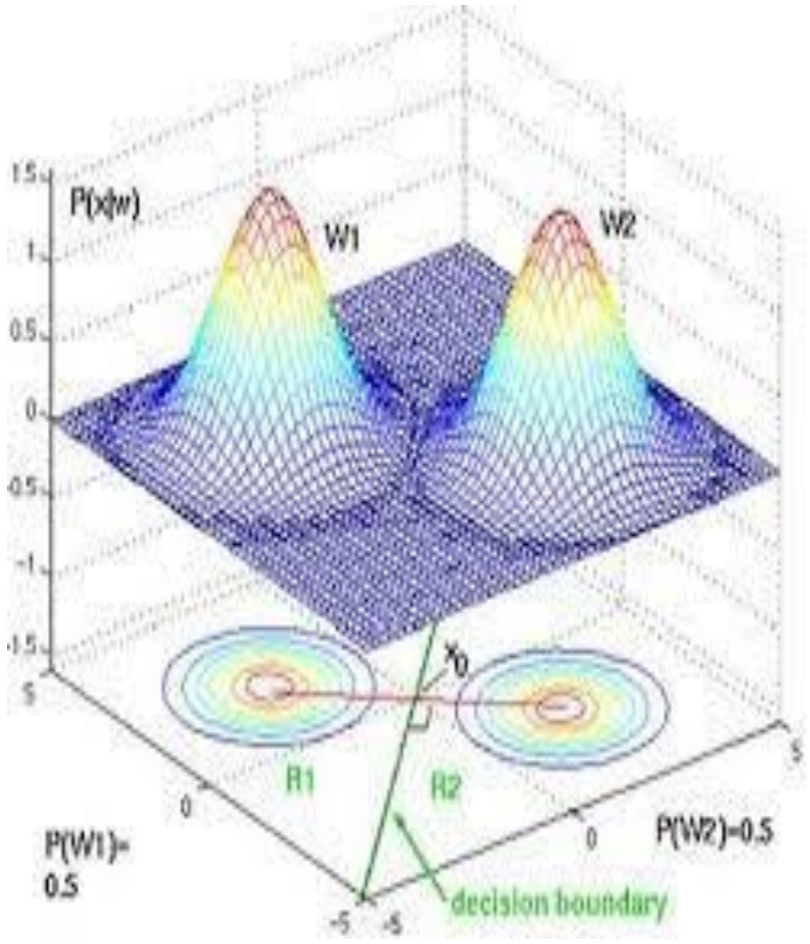
- ✓ Naive Bayes is used primarily to **predict the probability of different classes based on multiple attributes**.





Gaussian Naive Bayes

- ✓ Before, calculated the probabilities for input values for each class using a frequency. With real-valued inputs, we can calculate the mean and standard deviation of input values (x) for each class to summarize the distribution.
- ✓ This means that in addition to the probabilities for each class, we must also store the mean and standard deviations for each input variable for each class.
- ✓ Probabilities of new x values are calculated using the **Gaussian Probability Density Function (PDF).**
- ✓ When making predictions these parameters can be plugged into the Gaussian PDF with a new input for the variable, and in return the Gaussian PDF will provide an estimate of the probability of that new input value for that class.
- ✓ We can then plug in the probabilities into the equation above to make predictions with real-valued inputs.



Applications of Naive Bayes

- ✓ Classify a news article about technology, politics, or sports
- ✓ Sentiment analysis on social media
- ✓ Facial recognition software's
- ✓ Recommendation Systems as in Netflix, Amazon
- ✓ Spam filtering

References

- ✓ https://towardsdatascience.com/top-10-algorithms-for-machine-learning-beginners-149374935f3c

Francesco **Pugliese**