Naïve Bayes
vs
Gaussian Naïve Bayes

Yilin

Differences

1. Classical Naïve Bayes supports categorical features and modes each as conforming to a multinomial Distribution.
2. Gaussian Naïve Bayes supports continuous valued features and models each as conforming to a Gaussian (Normal) Distribution.
3. Categorical → Classical Naïve Bayes
4. Continuous → Gaussian Naïve Bayes

Note: If your data set contains both categorical and continuous features, you have two options.

- Discretize your continuous feature via bucketing or a similar methodology
- Hybrid Naïve Bayes method.

GAUSSIAN NAIVE BAYES CLASSIFIER

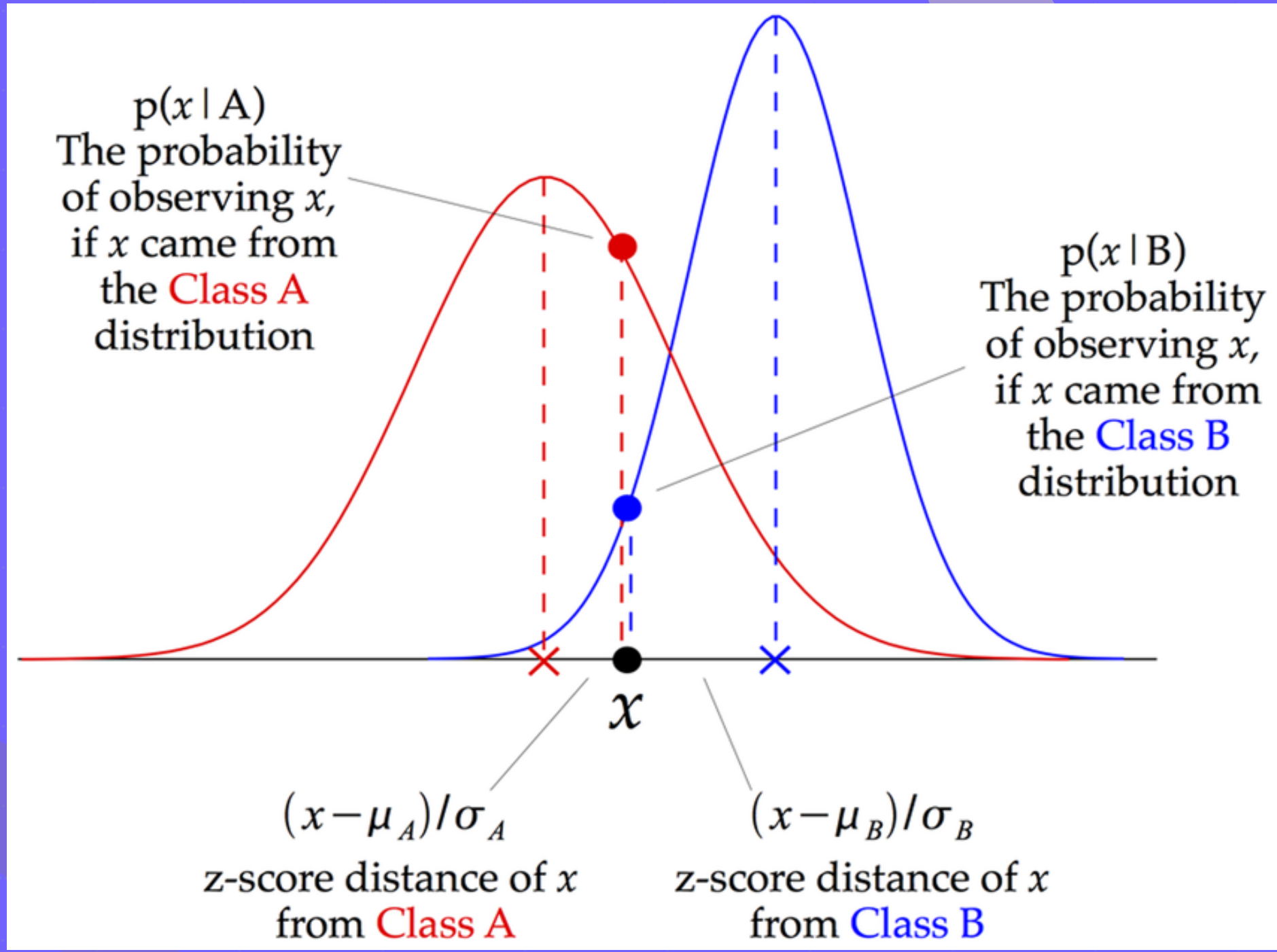
"Gaussian" because this is a normal distribution

This is our prior belief

$$p(\text{class} \mid \text{data}) = \frac{p(\text{data} \mid \text{class}) \times p(\text{class})}{p(\text{data})}$$

We don't calculate this in naive bayes classifiers

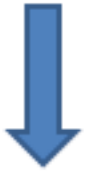
ChrisAlbon


p(x | A)
The probability of observing x, if x came from the Class A distribution
p(x | B)
The probability of observing x, if x came from the Class B distribution
x
$(x-\mu_A)/\sigma_A$
z-score distance of x from Class A
$(x-\mu_B)/\sigma_B$
z-score distance of x from Class B

Whether	Play
Sunny	No
Sunny	No
Overcast	Yes
Rainy	Yes
Rainy	Yes
Rainy	No
Overcast	Yes
Sunny	No
Sunny	Yes
Rainy	Yes
Sunny	Yes
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table

Whether	No	Yes
Overcast		4
Sunny	2	3
Rainy	3	2
Total	5	9



Likelihood Table 1

Whether	No	Yes		
Overcast		4	=4/14	0.29
Sunny	2	3	=5/14	0.36
Rainy	3	2	=5/14	0.36
Total	5	9		
	=5/14	=9/14		
	0.36	0.64		

Likelihood Table 2

Whether	No	Yes	Posterior Probability for No	Posterior Probability for Yes
Overcast		4	0/5=0	4/9=0.44
Sunny	2	3	2/5=0.4	3/9=0.33
Rainy	3	2	3/5=0.6	2/9=0.22
Total	5	9		

ID	GPA	GMAT	ADMITTED_IND
000000001	3.14	473	1
000000002	3.22	482	1
000000003	2.96	596	1
000000004	3.28	523	1
000000005	2.72	399	0
000000006	2.85	381	0
000000007	2.51	458	0
000000008	2.36	399	0

Which method?