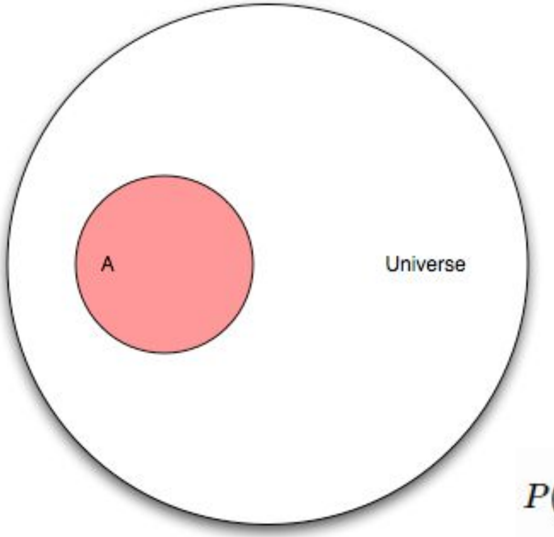
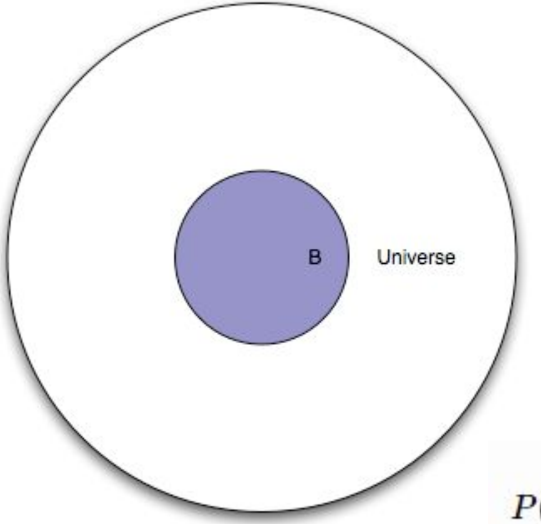
Lecture 5: Naive Bayes

Assaf Shocher

assafshocher@gmail.com

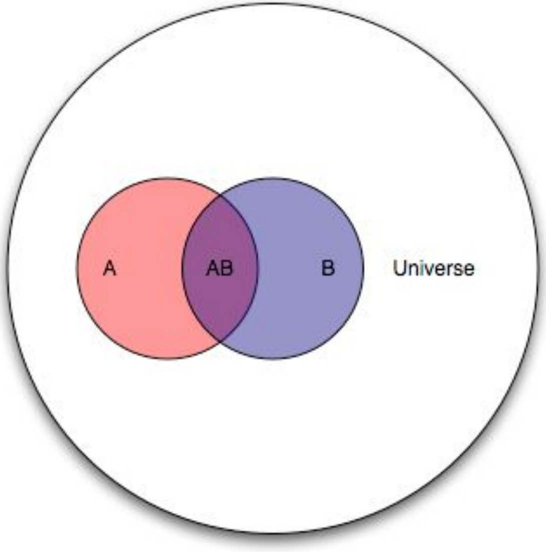


$$P(A) = \frac{|A|}{|U|}$$



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

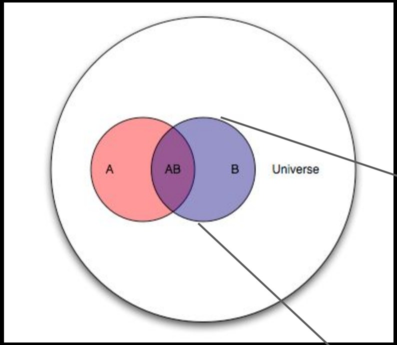
So what is the probability of A & B?



We can compute the probability of both events occurring (AB is a shorthand for $A \cap B$) in the same way.

$$P(AB) = \frac{|AB|}{|U|}$$

So what is the probability of A given B?



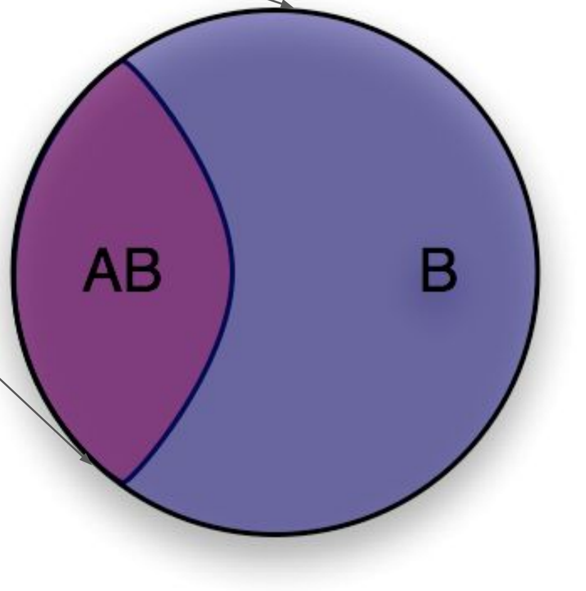
$$P(A|B) = \frac{|AB|}{|B|}$$

And if we divide both the numerator and the denominator by $|U|$

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

we can rewrite it using the previously derived equations as

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



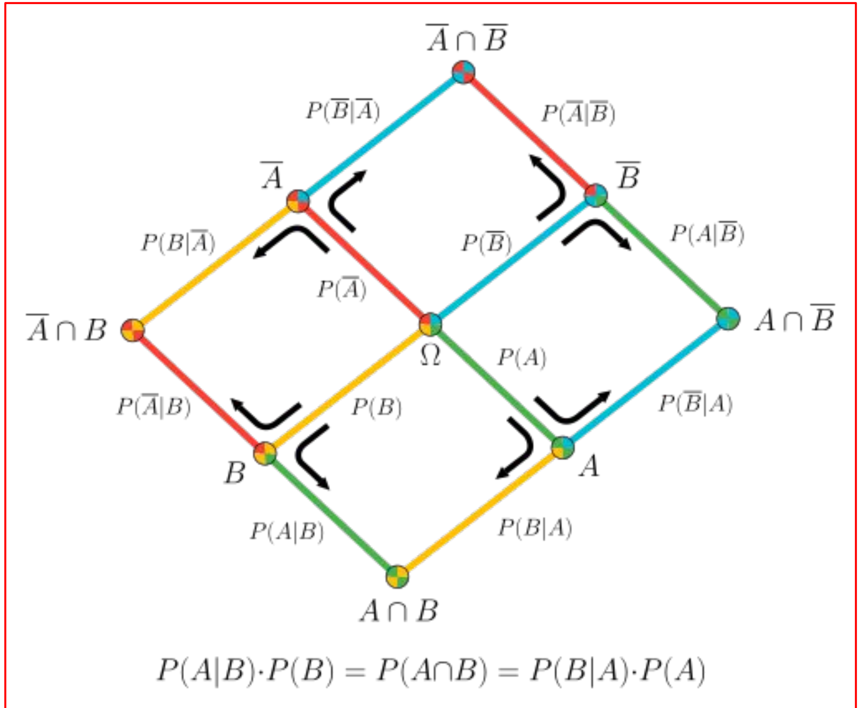
Reminder: Bayes' theorem

Likelihood / Evidence

Prior

Posterior

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$



Naive Bayes classifier

Is no more than applying Bayes' theorem, by estimating its terms.
We calculate the probability of belonging to Class k:

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$

$p(C_k) \approx$ (# of class k instances) / (total # of instances)

$p(x|C_k) \approx$ (# of x instances in class k) / (# of class k instances)

$p(x) \approx$ (# of x instances) / (total # of instances)

Example:

$$p(C_k) \approx (\#\text{ of class k instances}) \;/\; (\text{total }\#\text{ of instances})$$

$$p(x|C_k) \approx (\#\text{ of x instances in class k}) \;/\; (\#\text{ of class k instances})$$

$$p(x) \approx (\#\text{ of x instances}) \;/\; (\text{total }\#\text{ of instances})$$

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

More than one feature:

$$p(C_k|x_i) = p(x_i|C_k) \cdot p(C_k) / p(x_i)$$

$$p(C_k|x_1,x_2...x_N) = p(x_1,x_2...x_N|C_k) \cdot p(C_k) / p(x_1,x_2...x_N) =$$

$$[p(C_k) / p(x)] \cdot \Pi[p(x_i|C_k)]$$

So what is so naive about “Naive Bayes Classifier”???

So what is so naive about “Naive Bayes Classifier”???

$$p(C_k|x_i) = p(x_i|C_k) \cdot p(C_k) \,/\, p(x_i)$$

$$p(C_k|x_1,x_2 \ldots x_N) = p(x_1,x_2 \ldots x_N|C_k) \cdot p(C_k) \,/\, p(x_1,x_2 \ldots x_N) = [p(C_k) \,/\, p(x)] \cdot \Pi[p(x_i|C_k)]$$

What is the hidden assumption?

Problem: when data is continuous

What's the problem?

$p(C_k) \approx$ (# of class k instances) / (total # of instances)

$p(x|C_k) \approx$ (# of x instances in class k) / (# of class k instances)

$p(x) \approx$ (# of x instances) / (total # of instances)

# of x instances (in or not in class) is 0 with probability 1 !!!

Example (from wiki)

Is it a male or a female?

Gender	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

Gender	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

Solution: assume distribution (usually normal)

$$p(x = v \mid c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

Calculate mean and variance and you have it all!

Example solution

$$p(x = v \mid c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

Gender	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033e-02	176.25	1.2292e+02	11.25	9.1667e-01
female	5.4175	9.7225e-02	132.5	5.5833e+02	7.5	1.6667e+00

$$\text{posterior (male)} = \frac{P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male})}{evidence}$$

$$P(\text{male}) = 0.5$$

$$p(\text{height} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789,$$