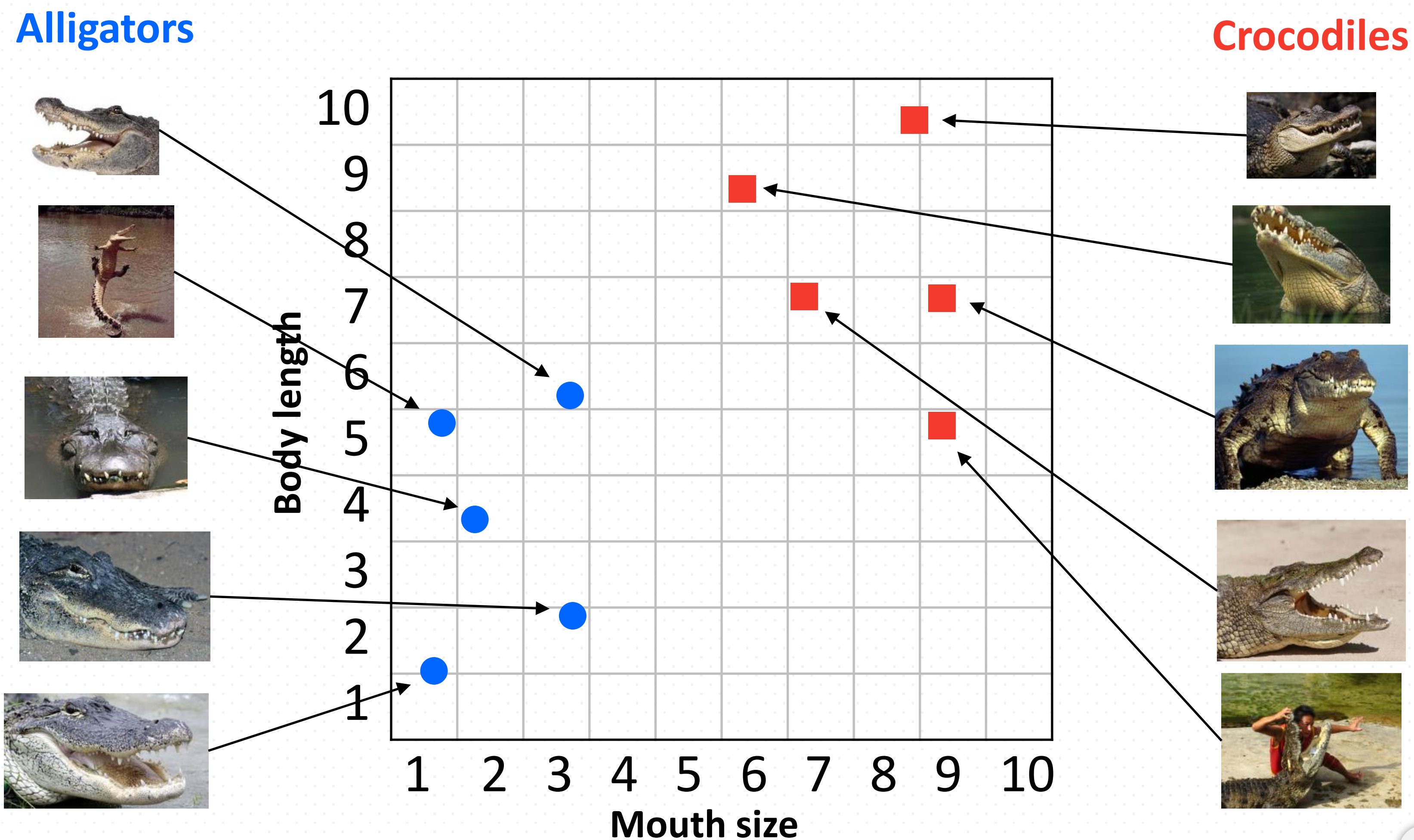
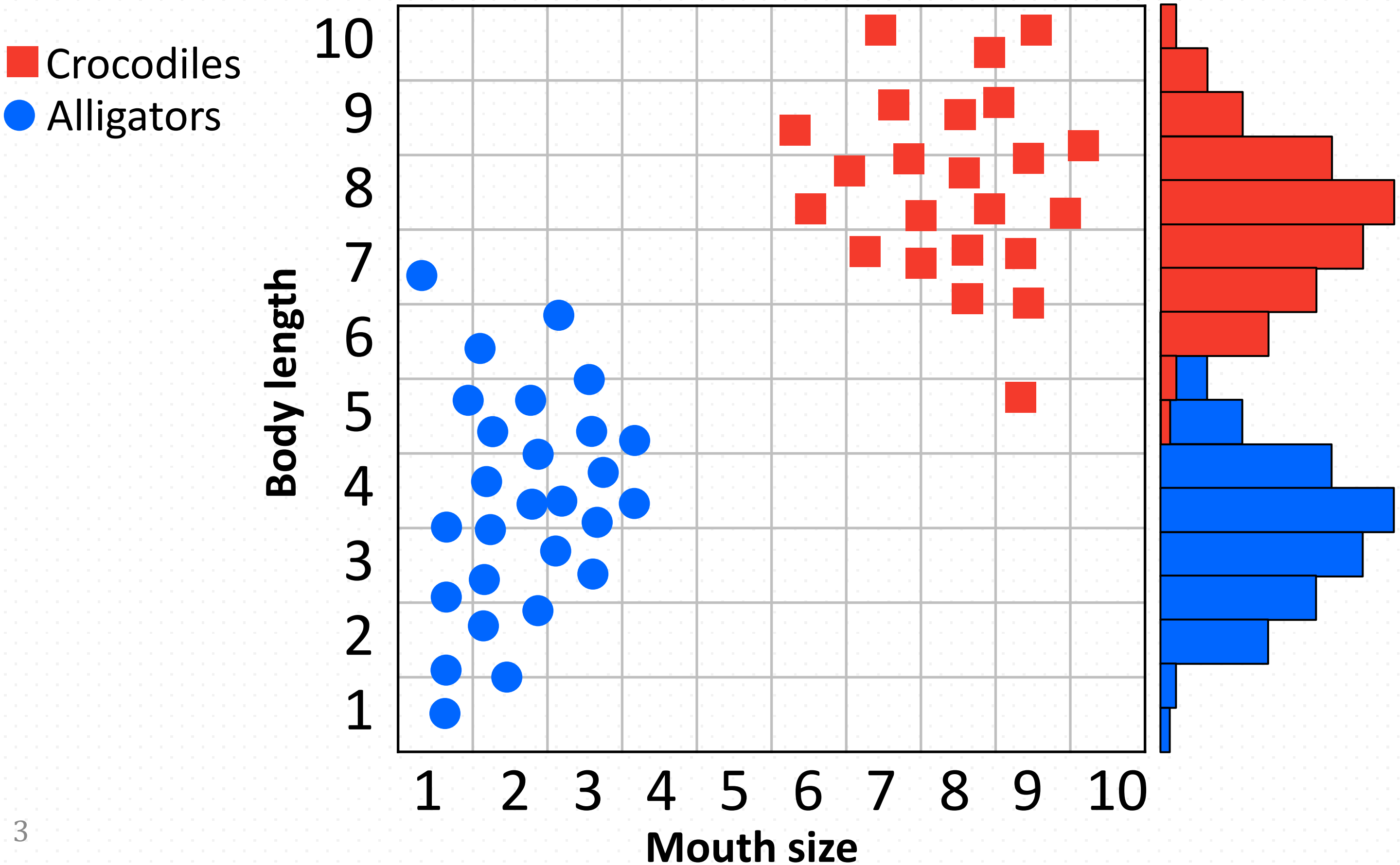
Naïve Bayes Classifier

(pages 231–238 on text book)

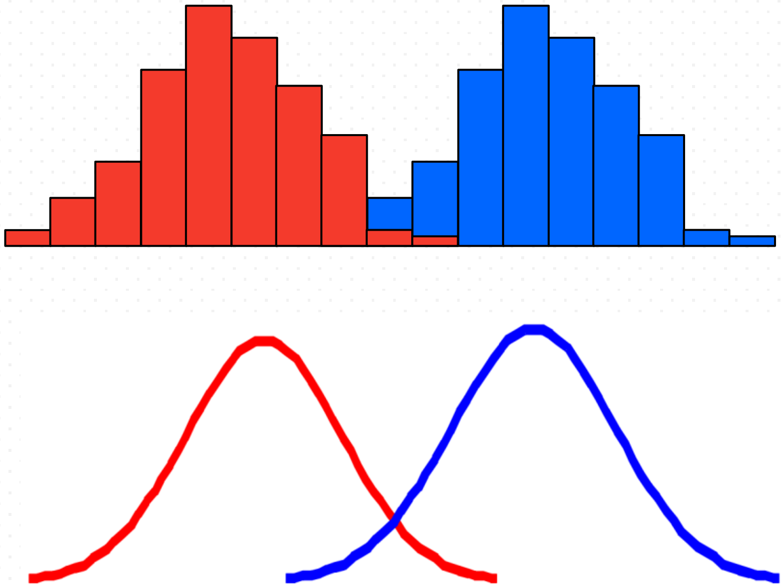
- Lets start off with a visual intuition
  - Adapted from Dr. Eamonn Keogh's lecture – UCR


Alligators
Crocodiles
10
9
8
7
6
5
4
3
2
1
Body length
1 2 3 4 5 6 7 8 9 10
Mouth size

Suppose we had a lot of data. We could use that data to build a histogram. Below is one built for the *body length* feature:

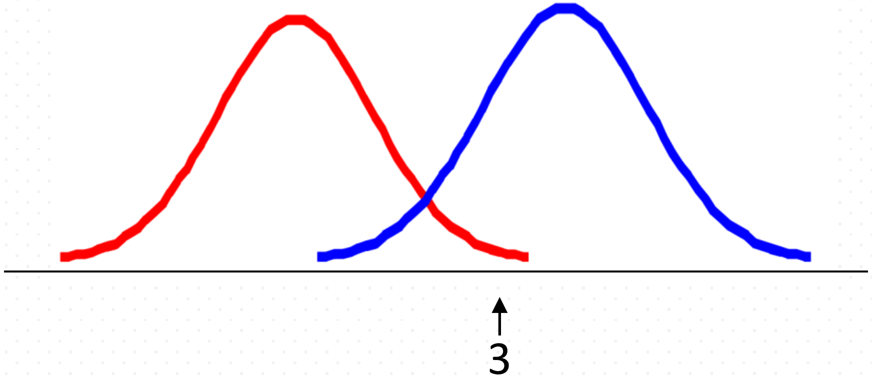


We can summarize these histograms as two normal distributions.



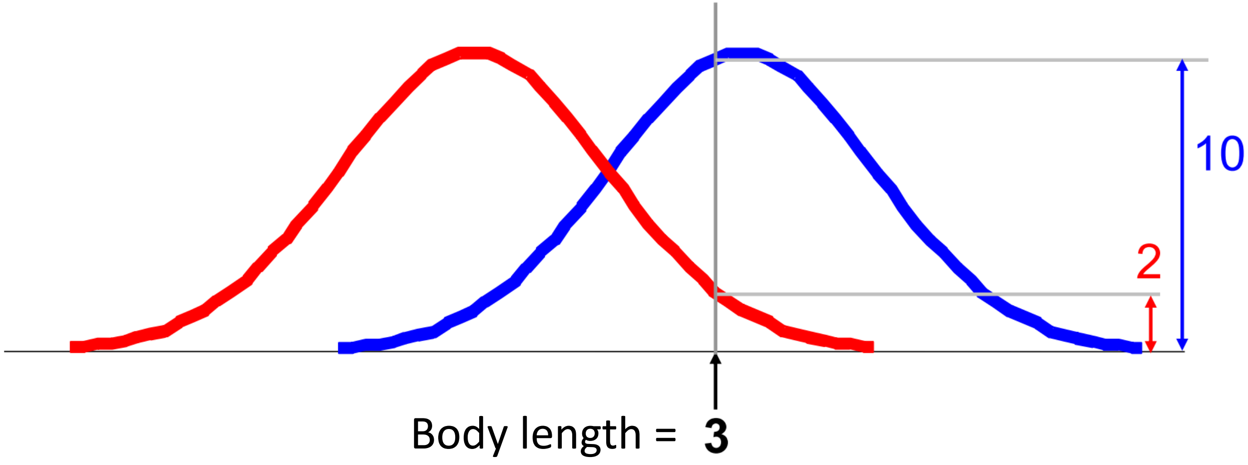
- Suppose we wish to classify a new animal that we just found. Its body length is 3 units. How can we classify it?
- One way to do this is, given the distributions of that feature, we can analyze which class is more *probable:* Crocodile or Alligator.
- More formally:

$$p(c_j|d) = probability\ of\ class\ \boldsymbol{c_j}, given\ that\ we\ observed\ \boldsymbol{d}$$



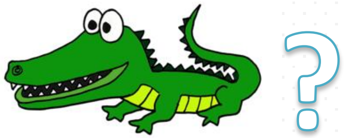
$$p(c_j|d) = probability\ of\ class\ \boldsymbol{c_j}, given\ that\ we\ observed\ \boldsymbol{d}$$

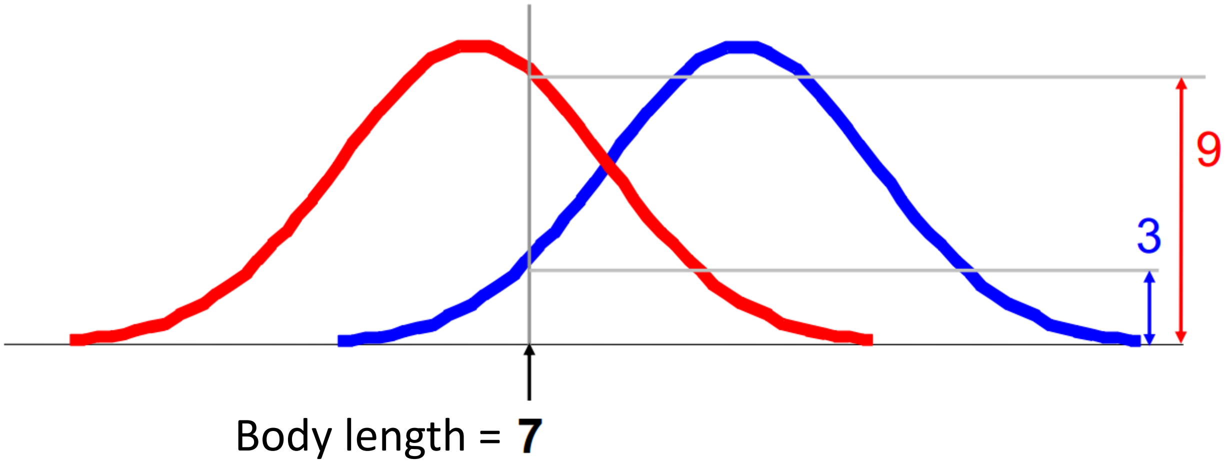
$$p(Alligator|body\ length = 3) = 10/(10+2) = \mathbf{0.833}$$
$$p(Crocodile|body\ length = 3) = 2/(10+2) = \mathbf{0.166}$$



$$p(c_j|d) = probability\ of\ class\ \boldsymbol{c_j}, given\ that\ we\ observed\ \boldsymbol{d}$$

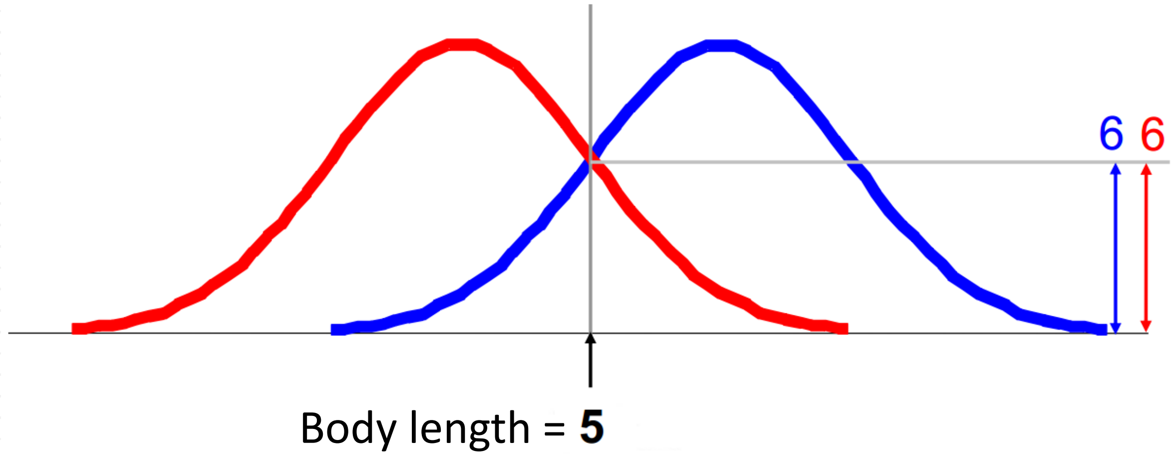
$$p(Alligator|body\ length = 7) = 3/(3+9) = \mathbf{0.25}$$
$$p(Crocodile|body\ length = 7) = 9/(3+9) = \mathbf{0.75}$$

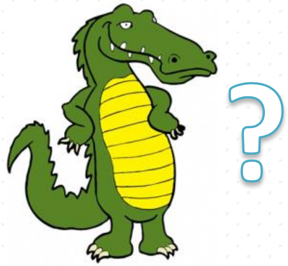




$$p(c_j|d) = probability\ of\ class\ \mathbf{c_j},\ given\ that\ we\ observed\ \boldsymbol{d}$$

$$p(Alligator|body\ length = 5) = 6/(6+6) = \mathbf{0.5}$$
$$p(Crocodile|body\ length = 5) = 6/(6+6) = \mathbf{0.5}$$





Naïve Bayes Classifier

- This visual intuition describes a simple Bayes classifier commonly known as:
  - Naïve Bayes
  - Simple Bayes
  - Idiot Bayes
- While going through the math, keep in mind the basic idea:

Given a new unseen instance, we (1) find its probability of it belonging to each class, and (2) pick the most probable.

Bayes Theorem

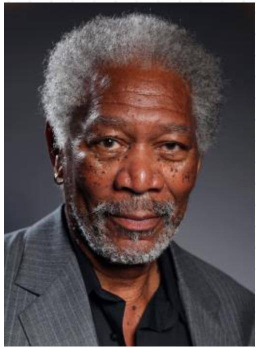
$$\underbrace{P(c_j|d)}_{\text{Posterior probability}} = \frac{\overbrace{P(d|c_j)}^{\text{Likelihood}}\overbrace{P(c_j)}^{\text{Class prior probability}}}{\underbrace{P(d)}_{\text{Predictor prior probability}}}$$

- $\boldsymbol{P(c_j|d)}$ = probability of instance d being in class c_j
- $\boldsymbol{P(d|c_j)}$ = probability of generating instance d given class c_j
- $\boldsymbol{P(c_j)}$ = probability of occurrence of class c_j
- $\boldsymbol{P(d)}$ = probability of instance d occurring

Suppose we have another binary classification problem with the following two classes: $c_1 = male$, and $c_2 = female$

We now have a person called *Morgan*. How do we classify them as male or female?

Morgan Fairchild



Morgan Freeman

What is the probability of being called Morgan given that you are a male?

What is the probability of being a male?

$$P(male|morgan) = \frac{P(morgan|male)P(male)}{P(morgan)}$$

What is the probability of being called Morgan

Suppose this individual on your left (Morgan) was arrested for money laundering. Is Morgan male or female?

Assume we are given the following database of names. We can then apply Bayes rule.

Name	Sex
Morgan	Male
Reid	Female
Morgan	Male
Morgan	Female
Everaldo	Male
Francis	Male
Jennifer	Female

$$P(c_j|d) = \frac{P(d|c_j)P(c_j)}{P(d)}$$

Name	Sex
Morgan	Male
Reid	Female
Morgan	Male
Morgan	Female
Everaldo	Male
Francis	Male
Jennifer	Female

$$P(female|morgan) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$P(male|morgan) = \frac{2/5 * 5/8}{3/8} = \frac{0.25}{3/8}$$

Money launderer Morgan is more likely to be male.

Naïve Bayes Classifier

- Both examples that we looked at considered only a single feature (i.e., *body length* and *name*)
- What if we have several features?

Name	Over 6ft	Eye color	Hair style	Sex
Morgan	Yes	Blue	Long	Female
Bob	No	Brown	None	Male
Vincent	Yes	Brown	Short	Male
Amanda	No	Brown	Short	Female
Reid	No	Blue	Short	Male
Lauren	No	Purple	Long	Female
Elisa	Yes	Brown	Long	Female

Naïve Bayes Classifier

- Naïve Bayes assumes that all features are independent (i.e., they have independent distributions).
- The probability of class c_j generating instance d can then be estimated as:

$$P(d|c_j) = P(d_1|c_j) \times P(d_2|c_j) \times \ldots \times P(d_n|c_j)$$

Probability of class c_j generating the observed value for feature 1

Probability of class c_j generating the observed value for feature 2

...

Naïve Bayes Classifier

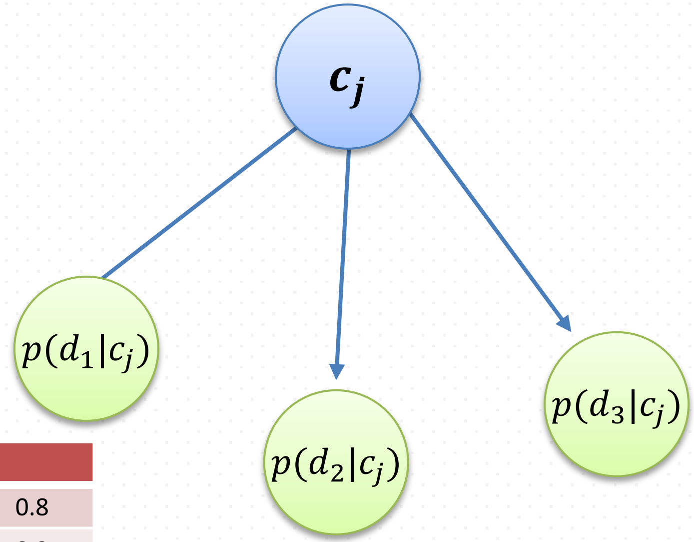
- Suppose we have Amanda's data:

Name	Over 6ft	Eye color	Hair style	Sex
Amanda	No	Brown	Short	?

$$\boldsymbol{P(Amanda|c_j)} = P(over6ft = No|c_j) \times P(eyecolor = Brown|c_j) \times P(hair = Short|c_j)$$

Naïve Bayes Classifier

- A visual representation:
  - Each class implies certain features with certain probabilities



All tables can be computed with a single scan of our dataset: **Very efficient!**

Sex	Over 6ft	
Male	Yes	0.8
	No	0.2
Female	Yes	0.01
	No	0.99

Sex	Eye color	
Male	Brown	0.67
	Blue	0.33
Female	Brown	0.45
	Blue	0.55

Sex	Hair	
Male	None	0.1
	Short	0.8
	Long	0.1
Female	None	0.05
	Short	0.15
	Long	0.8

Naïve Bayes Classifier

- Naïve Bayes is not sensitive to irrelevant features.
  - For instance, if we were trying to classify a person's sex based on a set of features that included *eye color* (which is irrelevant to their gender):

$$P(Reid|c_j) = P(eye = blue|c_j) \times P(wears\ makeup = Yes|c_j) \times \ \ldots$$

$$P(Reid|Female) = \boxed{9{,}000/10{,}000} \times 9{,}500/10{,}000 \times \ldots$$

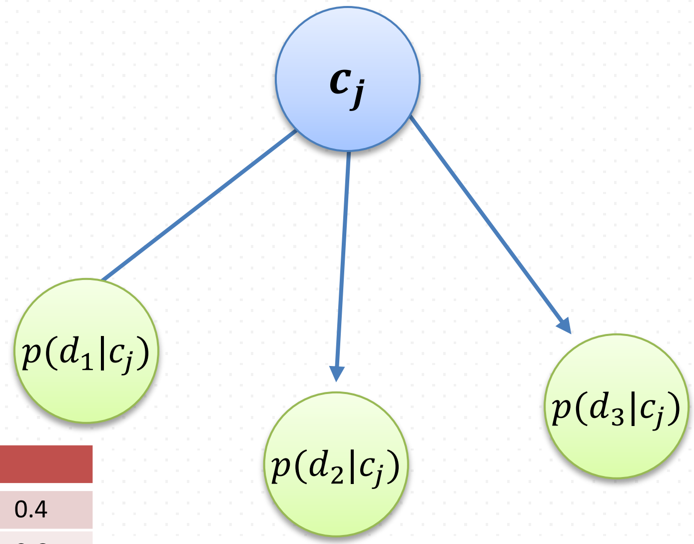
$$P(Reid|Male) = \boxed{9{,}007/10{,}000} \times 1/10{,}000 \times \ldots$$

Nearly identical!

  - The more data, the better!

Naïve Bayes Classifier

- Naïve Bayes works with any number of classes or features:



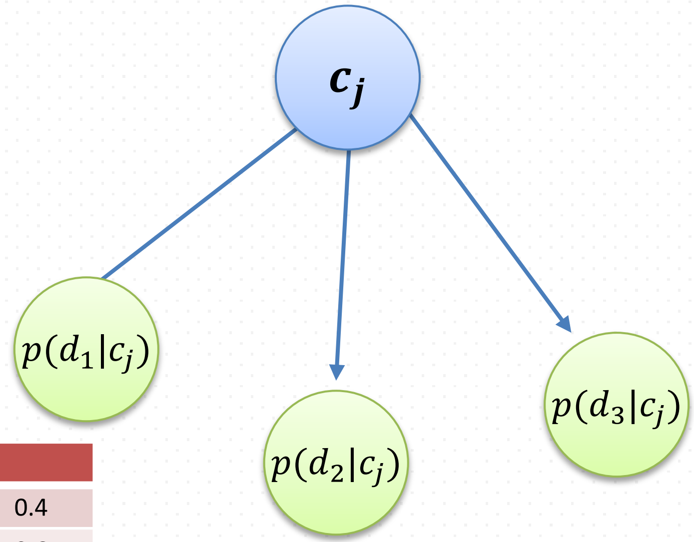
Sex	Friendly	
Crocodile	Yes	0.4
	No	0.6
Alligator	Yes	0.3
	No	0.7
T-Rex	Yes	0.05
	No	0.95

…

Sex	Color	
Crocodile	Black	0.1
	Brown	0.2
	Green	0.7
Alligator	Black	0.2
	Brown	0.3
	Green	0.5
T-Rex	Black	0.2
	Brown	0.2
	Green	0.6

Naïve Bayes Classifier - Problems

- Naïve Bayes assumes all features are independent (often **bad!**)



Sex	Friendly	
Crocodile	Yes	0.4
	No	0.6
Alligator	Yes	0.3
	No	0.7
T-Rex	Yes	0.05
	No	0.95

...

Sex	Color	
Crocodile	Black	0.1
	Brown	0.2
	Green	0.7
Alligator	Black	0.2
	Brown	0.3
	Green	0.5
T-Rex	Black	0.2
	Brown	0.2
	Green	0.6

Naïve Bayes Classifier - Problems

- What if one of the probabilities is zero? Ops.

$$P(d|c_j) = P(d_1|c_j) \times P(d_2|c_j) \times \ldots \times P(d_n|c_j)$$
$$= 0.15 \times \mathbf{0} \times \cdots \times 0.55$$

- This is not uncommon, especially when the number of instances is small
- Solution: **m-estimate**

m-estimate

- To avoid trouble when a probability $P(d_1|c_j) = 0$, we fix its prior probability and the number of samples to some non-zero value beforehand
  - Think of it as adding a bunch of fake instances before we start the whole process
- If we create $m > 0$ fake samples of feature X with value of x, and we assign a prior probability p to them, then posterior probabilities are obtained as:

$$P(X = x|c_j) = \frac{\#(X=x, c_j)+mp}{\#(c_j)+m}$$

Naïve Bayes Pros

- Can be used for both continuous and discrete features.
- Can be used for binary or multi-class problems.
- It is robust to isolated noise/outliers.
- Missing values are easily handled. They are simply ignored.
- It is robust to irrelevant attributes.

Naïve Bayes Cons

- Correlated features degrade performance. If we have the same feature multiple times in the dataset, we can change classification results.
- When m-estimate is used, the choice of *m's* and *p's* is very arbitrary.

Naïve Bayes – Class exercise

- Predict if Bob will default his loan

<u>Bob</u>
Home owner: *No*
Marital status: *Married*
Job experience: *3*

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

Naïve Bayes – Class exercise (1)

Bob

Home owner: *No*
Marital status: *Married*
Job experience: *3*

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

- $P(Y = No) = 7/10$
- $P(Home\ owner = No|Y = No) = 4/7$
- $P(Marital\ status = Married|Y = No) = 4/7$
- $P(Job\ experience = 3|Y = No) = 2/7$

$$P(Bob\ will\ NOT\ default) = \frac{7}{10} \times \frac{4}{7} \times \frac{4}{7} \times \frac{2}{7} = \mathbf{0.065}$$

Naïve Bayes – Class exercise (2)

Bob

Home owner: *No*
Marital status: *Married*
Job experience: *3*

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes

- $P(Y = Yes) = 3/10$
- $P(Home\ owner = No|Y = Yes) = 1/3$
- $P(Marital\ status = Married|Y = Yes) = 1/3$
- $P(Job\ experience = 3|Y = Yes) = 1/3$

$$P(Bob\ will\ default) = \frac{3}{10} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \mathbf{0.011}$$

Naïve Bayes – Class exercise (3)

Bob

Home owner: *No*
Marital status: *Married*
Job experience: *3*

- $P(Bob\ will\ NOT\ default) = \mathbf{0.065}$
- $P(Bob\ will\ default) = \mathbf{0.011}$

Predict: BOB WILL NOT DEFAULT

Home owner	Marital Status	Job experience (1-5)	Defaulted
Yes	Single	3	No
No	Married	4	No
No	Single	5	No
Yes	Married	4	No
No	Divorced	2	Yes
No	Married	4	No
Yes	Divorced	2	No
No	Married	3	Yes
No	Married	3	No
Yes	Single	2	Yes