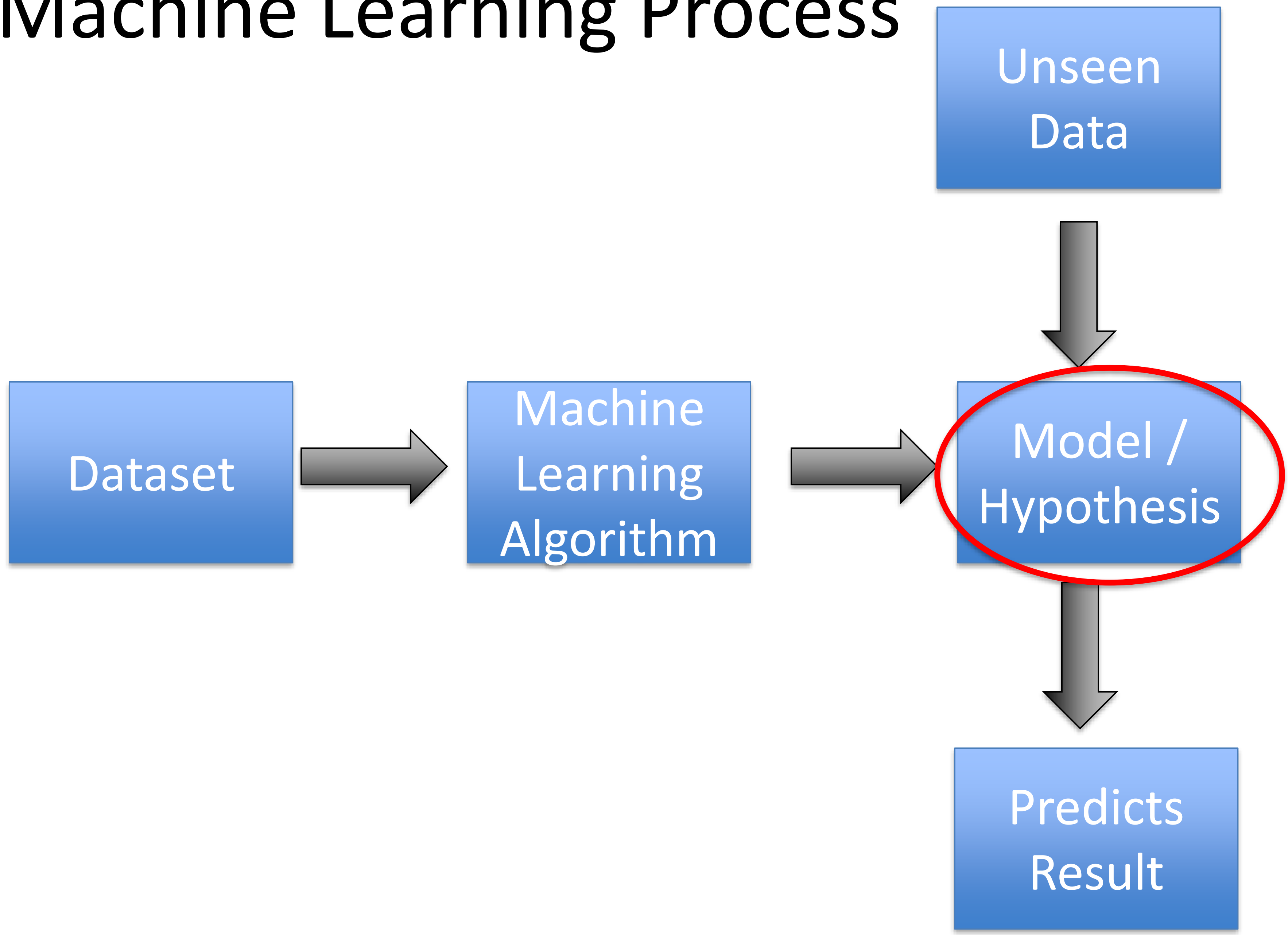
Machine Learning

Machine Learning

Lecture: Bayesian Classification

Ted Scully

Machine Learning Process



Training a Naïve Bayes Classifier

- We estimate conditional probabilities for each arc from the training data!
- Simple frequency estimates:
- $P(X = x_1 | C = c_1) = \frac{N_{x1c1}}{N_{c1}}$
  - N_{x1c1} = counts of cases where $X=x_1$ and $C=c_1$
  - N_{c1} = count of cases where $C=c_1$

P(name = "Joe" | class = male) =

How many of male class are called Joe / Total number male class

Naïve Bayes Classifier – Worked Example

- We are now going to step through a worked example with the tennis dataset below.

  - **Step 1**: Generate all conditional probabilities

  - **Step 2**: Perform inference by passing our model the following feature vector
    <Outlook =sunny, Temp = cold, Humidity = high, Windy = true>

Anyone for Tennis?

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

Anyone for Tennis?

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

Naïve Bayes
ML Algorithm

Conditional Probabilities

1. P(Outlook = Sunny | Play = Y)
2. P(Outlook = Sunny | Play = N)
3. P(Outlook = Overcast | Play = Y)
4. P(Outlook = Overcast | Play = N)
5. .....

Conditional Probabilities for Play = Y

Anyone for Tennis?				
ID	Outlook	Temp	Humidity	Windy
A	sunny	hot	high	false
B	sunny	hot	high	true
C	overcast	hot	high	false
D	rainy	mild	high	false
E	rainy	cool	normal	false
F	rainy	cool	normal	true
G	overcast	cool	normal	true
H	sunny	mild	high	false
I	sunny	cool	normal	false
J	rainy	mild	normal	false
K	sunny	mild	normal	true
L	overcast	mild	high	true
M	overcast	hot	normal	false
N	rainy	mild	high	true

P(Out = Sunny | Play = Y)

P(Out = Overcast | Play = Y)

P(Out = Rainy | Play = Y)

P(Temp = Hot | Play = Y)

P(Temp = Mild | Play = Y)

P(Temp = Cool | Play = Y)

P(Hum= High | Play = Y)

P(Hum= Normal | Play = Y)

P(Win= False | Play = Y)

P(Win= True | Play = Y)

Conditional Probabilities for Play = Y

Anyone for Tennis?				
ID	Outlook	Temp	Humidity	Windy
A	sunny	hot	high	false
B	sunny	hot	high	true
C	overcast	hot	high	false
D	rainy	mild	high	false
E	rainy	cool	normal	false
F	rainy	cool	normal	true
G	overcast	cool	normal	true
H	sunny	mild	high	false
I	sunny	cool	normal	false
J	rainy	mild	normal	false
K	sunny	mild	normal	true
L	overcast	mild	high	true
M	overcast	hot	normal	false
N	rainy	mild	high	true

Conditional Probabilities also need to be worked out for Play = N

P(Out = Sunny | Play = N)

P(Out = Overcast | Play = N)

P(Out = Rainy | Play = N)

P(Temp = Hot | Play = N)

P(Temp = Mild | Play = N)

P(Temp = Cool | Play = N)

P(Hum= High | Play = N)

P(Hum= Normal | Play = N)

P(Win= False | Play = N)

P(Win= True | Play = N)

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

- We will first look at calculating the probabilities needed for the class 'play' and the feature '**windy**'. Lets work out the **P(Wind = t | Play = y)**

- $P(X = x_1 | C = c_1) = \frac{N_{x1c1}}{N_{c1}}$
  - N_{x1c1} = counts of cases where $X=x_1$ and $C=c_1$
  - N_{c1} = count of cases where $C=c_1$

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

- We will first look at calculating the probabilities needed for the class 'play' and the attribute '**windy**'. Lets work out the P(Wind = t | Play = y)

As we can see from the image, there are 9 cases where **`Play=y`**. In 3 of these, **`Wind=t`**, and in the other 6, **`Wind=f`**. Therefore, the probability of **`Wind=t`** given that **`Play=y`** is 3/9, according to these observations.

```
P(Wind=t | Play=y) = 3/9
P(Wind=f | Play=y) = 6/9
```

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

Next work out **P(Wind = t | Play = n)** and **P(Wind = f | Play = n)**

As we can see from the image, there are 5 cases where `Play=n`. In 3 of these, `Wind=t`, and in the other 2, `Wind=f`. Therefore:

```
P(Wind=t | Play=n) = 3/5
P(Wind=f | Play=n) = 2/5
```

We now have all four probabilities we need for the arc between play and windy. Next we need to apply the same method to calculate the probabilities for each of the other arcs.

Our Naïve Bayes algorithm takes as input the data set and produces the following **model.**

```
P(Outlook=s | Play=y) = 2/9  P(Outlook=s | Play=n) = 3/5
P(Outlook=o | Play=y) = 4/9  P(Outlook=o | Play=n) = 0/5
P(Outlook=r  | Play=y) = 3/9 P(Outlook=r  | Play=n) = 2/5
```

```
P(Wind=t | Play=y) = 3/9   P(Wind=t | Play=n) = 3/5
P(Wind=f | Play=y) = 6/9   P(Wind=f | Play=n) = 2/5
```

```
P(Temp=h | Play=y) = 2/9  P(Temp=h  | Play=n) = 2/5
P(Temp=m | Play=y) = 4/9  P(Temp=m | Play=n) = 2/5
P(Temp=c  | Play=y) = 3/9 P(Temp=c  | Play=n) = 1/5
```

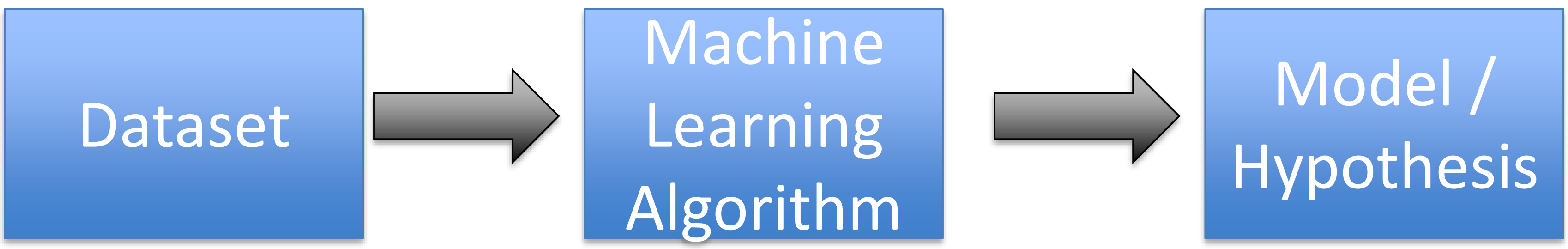
P(Humidity =high| Play = yes) = 3/9
P(Humidity =normal| Play = yes) = 6/9

P(Humidity =high| Play = no) = 4/5
P(Humidity =normal| Play = no) = 1/5

```
Play=y) = 9/14
Play=n) = 5/14
```

Machine Learning Process

We have calculated all the conditional probabilities. Therefore, we have now built the model



Naïve Bayes Classifier – Worked Example

- We are now going to step through a worked example with the tennis dataset below.
  - **Step 1**: Generate all conditional probabilities
  - **Step 2**: Perform inference by passing our model the following feature vector
    <Outlook =<u>sunny</u>, Temp = <u>cold</u>, Humidity = <u>high</u>, Windy = <u>true</u>>

Anyone for Tennis?

ID	Outlook	Temp	Humidity	Windy	Play?
A	sunny	hot	high	false	no
B	sunny	hot	high	true	no
C	overcast	hot	high	false	yes
D	rainy	mild	high	false	yes
E	rainy	cool	normal	false	yes
F	rainy	cool	normal	true	no
G	overcast	cool	normal	true	yes
H	sunny	mild	high	false	no
I	sunny	cool	normal	false	yes
J	rainy	mild	normal	false	yes
K	sunny	mild	normal	true	yes
L	overcast	mild	high	true	yes
M	overcast	hot	normal	false	yes
N	rainy	mild	high	true	no

Classify a New Instance

$$\underset{c \in C}{argmax} \left(\prod_{x \in X} P(x|c) \;\; P(c) \right)$$

- *Will I play tennis under the following conditions:*

<Outlook =sunny, Temp = cold, Humidity = high, Windy = true>

Play is y or n. Evaluate probability of each given data.

P(Play = y | Outlook = s, Temp =c, Humidity = h, Wind = t) =

P(Play = n | Outlook = s, Temp =c, Humidity = h, Wind = t) =

Classify a New Instance

$$\underset{c \in C}{argmax}\left(\prod_{x \in X} P(x|c) \;\; P(c)\right)$$

- *Will I play tennis under the following conditions:*

<Outlook =<u>sunny</u>, Temp = <u>cold</u>, Humidity = <u>high</u>, Windy = <u>true></u>

Play is y or n. Evaluate probability of each given data.

P(Play = y | Outlook = s, Temp =c, Humidity = h, Wind = t) = P(Outlook = s | Play = y) * P(Temp=c | Play = y) * P(Humidity= h | Play = y) * P(Wind = t | Play = y) * P(Play = y)

P(Play = n | Outlook = s, Temp =c, Humidity = h, Wind = t) = P(Outlook = s | Play = n) * P(Temp=c | Play = n) * P(Humidity= h | Play = n) * P(Wind = t | Play = n) * P(Play = n)

Classify a New Instance

$$\underset{c \in C}{argmax} \left(\prod_{x \in X} P(x|c) \;\; P(c) \right)$$

- *Will I play tennis under the following conditions:*

<Outlook =sunny, Temp = cold, Humidity = high, Windy = true>

Play is y or n. Evaluate probability of each given data.

P(Play = y | Outlook = s, Temp =c, Humidity = h, Wind = t) =
P(Outlook = s | Play = y) * P(Temp=c | Play = y) * P(Humidity= h | Play = y) * P(Wind = t | Play = y) * P(Play = y)
= 2/9 * 3/9 * 3/9 * 3/9* 9/14 = **0.005291**

P(Play = n | Outlook = s, Temp =c, Humidity = h, Wind = t) =
P(Outlook = s | Play = n) * P(Temp=c | Play = n) * P(Humidity= h | Play = n) * P(Wind = t | Play = n) * P(Play = n)
= 3/5 * 1/5 * 4/5 *3/5 * 5/14 = **0.020571**

Normalise the Results

$$\underset{c \in C}{argmax}\left(\prod_{x \in X} P(x|c) \quad P(c)\right)$$

P(Play = y | data) = 0.005291
P(Play = n | data) = 0.020571

(Why do above probabilities not add to 1?)

P(Play = y | data) = (0.005291*100) / (0.005291+ 0.020571) = 20.5%
P(Play = n | data) = (0.020571 *100) / (0.005291+ 0.020571) = 79.5%

Conclusion: more likely NOT to play tennis today.

From the calculations, it is seen that the probability of `Play=yes` is `20.5%`, whereas the probability of `Play=no` is `79.5%`. Selecting the outcome with the higher probability, the classification is that Play=no.

$$\underset{c \in C}{argmax} \left(\prod_{x \in X} P(x|c) \quad P(c) \right)$$

Consider the following data instance:

Outlook =**overcast**, Temp = mild, Humidity = normal, Windy = false: Play = ?

```
P(Outlook=s | Play=y) = 2/9  P(Outlook=s | Play=n) = 3/5
P(Outlook=o | Play=y) = 4/9  P(Outlook=o | Play=n) = 0/5
P(Outlook=r  | Play=y) = 3/9 P(Outlook=r  | Play=n) = 2/5
```

```
P(Wind=t | Play=y) = 3/9   P(Wind=t | Play=n) = 3/5
P(Wind=f | Play=y) = 6/9   P(Wind=f | Play=n) = 2/5
```

```
P(Temp=h | Play=y) = 2/9  P(Temp=h  | Play=n) = 2/5
P(Temp=m | Play=y) = 4/9  P(Temp=m | Play=n) = 2/5
P(Temp=c  | Play=y) = 3/9 P(Temp=c  | Play=n) = 1/5
```

P(Humidity =high| Play = yes) = 3/9
P(Humidity =normal| Play = yes) = 6/9

P(Humidity =high| Play = no) = 4/5
P(Humidity =normal| Play = no) = 1/5

```
Play=y) = 9/14
Play=n) = 5/14
```

$$\underset{c \in C}{argmax}\left(\prod_{x \in X} P(x|c) \quad P(c)\right)$$

Consider the following data instance:

Outlook =<u>overcast</u>, Temp = mild, Humidity = normal, Windy = false: Play = ?

```
P(Outlook=s | Play=y) = 2/9  P(Outlook=s | Play=n) = 3/5
P(Outlook=o | Play=y) = 4/9  P(Outlook=o | Play=n) = 0/5
P(Outlook=r  | Play=y) = 3/9 P(Outlook=r  | Play=n) = 2/5
```

Problem with Using Frequencies for Probability Calculations

- So far we estimated probabilities using the following:

- $P(X = x_1 | C = c_1) = \frac{\boldsymbol{N_{x1c1}}}{\boldsymbol{N_{c1}}}$
  - N_{x1c1} = counts of cases where $X=x_1$ and $C=c_1$
  - N_{c1} = count of cases where $C=c_1$

- To avoid the problem of <u>zero probabilities</u> we can applying basic smoothing techniques to the above formula.

Avoiding Zeros

- To avoid the problem outlined on the previous slide we typically use +1 or laplace smoothing.

- Often some basic softening of the equation is performed. For example (**+1 smoothing**)
- $P(X = x_1 | C = c_1) = \frac{\boldsymbol{N_{x1c1}+1}}{\boldsymbol{N_{c1}+2}}$

- ***Laplace* Smoothing** (m-estimate) :
- $P(X = x_1 | C = c_1) = \frac{\boldsymbol{N_{x1c1}+1}}{\boldsymbol{N_{c1+|X|}}}$

    - N_{x1c1} = counts of cases where X=x1 and C=c1
    - N_{c1} = count of cases where C=c1
    - |X| = count of cases of X (number of features in the dataset)

Avoiding Zeros

- Remember we worked out **P(Outlook = o | Play = n) = 0/5**

- +1 smoothing $(N_{x1c1} + 1) / (N_{c1} + 2)$
- If we use +1 smoothing **P(Outlook = o | Play = n) would be (0+1)/(5+2) = 1/7**

- *Laplace* Smoothing (m-estimate) : $(N_{x1c1} + 1) / (N_{c1} + |X|)$
- **P(Outlook = o | Play = n) would be (0+1)/(5+4) = 1/9**

  - Remember |X| is the number of features

Problems with Probabilities for Naïve Bayes

$$\underset{c \in C}{argmax} \left(\prod_{x \in X} P(x|c) \;\; P(c) \right)$$

Can you see any **computational problem** that may occur from this formula? Hint: What might happen if you have a large amount of features?

The computation issue is that of underflow: doing too many multiplications of small numbers.

When we go to calculate the product p(w0|ci)p(w1|ci)p(w2|ci)...p(wN|ci) and many of these numbers are very small, we'll get underflow (multiply many small numbers in a programming language and eventually it rounds off to 0.)

Using Log

- The most common solution to the problem on the previous slide is to calculate the **logarithm of this product**.
- Doing this allows us to avoid the underflow or round-off error problem. Why? Because we end up adding the individual probabilities rather than multiplying them
- In other word we now get the log of the Bayes equation

$$\mathbf{log}\Big(\prod_{x \in X} P(x|c)\, P(c)\Big)$$

  - We now use

$$\sum_{x \in X} \boldsymbol{log} P(x|c) \; + \boldsymbol{log} P(c)$$