Ollscoil Teicneolaíochta na Mumhan
Munster Technological University

Machine Learning

Lecture 9: Naïve Bayesian classification

(Naïve) likelihood-ratio test

- The optimal Bayesian decision rule for minimum-error-rate classification is to threshold the **likelihood-ratio** as follows

$$\frac{P[x_1, \ldots, x_n|\omega_1]}{P[x_1, \ldots, x_n|\omega_2]} > \frac{P[\omega_2]}{P[\omega_1]}$$

- If we make the (naïve) assumption that all features $x_1, \ldots, x_n$ are statistically independent, then this is equivalent to multiplying the likelihoods of all features per class and evaluate

$$\frac{\prod_{i=1}^{n} P[x_i|\omega_1]}{\prod_{i=1}^{n} P[x_i|\omega_2]} > \frac{P[\omega_2]}{P[\omega_1]}$$

(Naïve) likelihood-ratio test

- Therefore, all we need to calculate are the priors for each class

$$P[\omega_1], P[\omega_2]$$

- And all likelihoods per feature per class

$$P[x_1|\omega_1], \ldots, P[x_n|\omega_1]$$

and

$$P[x_1|\omega_2], \ldots, P[x_n|\omega_2]$$

Example: Tennis dataset

The number of samples are $\#samples = 14$

The two classes are

ω_1: Play=yes

ω_2: Play=no

The number of samples per class are

$$\#[Play = yes] = 9$$
$$\#[Play = no] = 5$$

The priors then are

$$P[\omega_1] = \frac{\#[Play = yes]}{\#samples} = \frac{9}{14} = .64$$

$$P[\omega_2] = \frac{\#[Play = no]}{\#samples} = \frac{5}{14} = .36$$

Note, how the probabilities add up to 1.

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

The four features are

x_1: Outlook
x_2: Temp
x_3: Humidity
x_4: Windy

The number of samples with feature x_1 in class ω_1 are

$$\#[Outlook = sunny | Play = yes] = 2$$
$$\#[Outlook = overcast | Play = yes] = 4$$
$$\#[Outlook = rainy | Play = yes] = 3$$

Therefore the likelihood $P[x_1|\omega_1]$ is given by

$$P[sunny|\omega_1] = \frac{\#[Outlook = sunny|Play = yes]}{\#[Play = yes]} = \frac{2}{9} = .22$$
$$P[overcast|\omega_1] = \frac{\#[Outlook = overcast|Play = yes]}{\#[Play = yes]} = \frac{4}{9} = .44$$
$$P[rainy|\omega_1] = \frac{\#[Outlook = rainy|Play = yes]}{\#[Play = yes]} = \frac{3}{9} = .33$$

Note, how the probabilities add up to 1.

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

We do the same for the class ω_2

Then the number of samples are

$$\#[Outlook = sunny|Play = no] = 3$$
$$\#[Outlook = overcast|Play = no] = 0$$
$$\#[Outlook = rainy|Play = no] = 2$$

Therefore the likelihood $P[x_1|\omega_2]$ is

$$P[sunny|\omega_2] = \frac{\#[Outlook = sunny|Play = no]}{\#[Play = no]} = \frac{3}{5} = .6$$
$$P[overcast|\omega_2] = \frac{\#[Outlook = overcast|Play = no]}{\#[Play = no]} = \frac{0}{5} = 0$$
$$P[rainy|\omega_2] = \frac{\#[Outlook = rainy|Play = no]}{\#[Play = no]} = \frac{2}{5} = .4$$

Note, that $P[overcast|\omega_2] = 0$. We will see later, how that is an issue and how to overcome this issue.

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

The number of samples are for feature x_2 are

$$\#[Temp = hot|Play = yes] = 2$$
$$\#[Temp = mild|Play = yes] = 4$$
$$\#[Temp = cool|Play = yes] = 3$$
$$\#[Temp = hot|Play = no] = 2$$
$$\#[Temp = mild|Play = no] = 2$$
$$\#[Temp = cool|Play = no] = 1$$

Therefore the likelihoods $P[x_2|\omega_1]$ and $P[x_2|\omega_2]$ are

$$P[hot|\omega_1] = \frac{\#[Temp = hot|Play = yes]}{\#[Play = yes]} = \frac{2}{9} = .22$$
$$P[mild|\omega_1] = \frac{\#[Temp = mild|Play = yes]}{\#[Play = yes]} = \frac{4}{9} = .44$$
$$P[cool|\omega_1] = \frac{\#[Temp = cool|Play = yes]}{\#[Play = yes]} = \frac{3}{9} = .33$$
$$P[hot|\omega_2] = \frac{\#[Temp = hot|Play = no]}{\#[Play = no]} = \frac{2}{5} = .4$$
$$P[mild|\omega_2] = \frac{\#[Temp = mild|Play = no]}{\#[Play = no]} = \frac{2}{5} = .4$$
$$P[cool|\omega_2] = \frac{\#[Temp = cool|Play = no]}{\#[Play = no]} = \frac{1}{5} = .2$$

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

The number of samples are for feature x_3 are

$$\#[Humidity = high|Play = yes] = 3$$
$$\#[Humidity = normal|Play = yes] = 6$$
$$\#[Humidity = high|Play = no] = 4$$
$$\#[Humidity = normal|Play = no] = 1$$

Therefore the likelihoods $P[x_3|\omega_1]$ and $P[x_3|\omega_2]$ are

$$P[high|\omega_1] = \frac{\#[Humidity = high|Play = yes]}{\#[Play = yes]} = \frac{3}{9} = .33$$
$$P[normal|\omega_1] = \frac{\#[Humidity = normal|Play = yes]}{\#[Play = yes]} = \frac{6}{9} = .66$$

$$P[high|\omega_2] = \frac{\#[Humidity = high|Play = no]}{\#[Play = no]} = \frac{4}{5} = .8$$
$$P[normal|\omega_2] = \frac{\#[Humidity = normal|Play = no]}{\#[Play = no]} = \frac{1}{5} = .2$$

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

The number of samples are for feature x_4 are

$$\#[Windy = yes|Play = yes] = 3$$
$$\#[Windy = no|Play = yes] = 6$$
$$\#[Windy = yes|Play = no] = 3$$
$$\#[Windy = no|Play = no] = 2$$

Therefore the likelihoods $P[x_4|\omega_1]$ and $P[x_4|\omega_2]$ are

$$P[yes|\omega_1] = \frac{\#[Windy = yes|Play = yes]}{\#[Play = yes]} = \frac{3}{9} = .33$$
$$P[no|\omega_1] = \frac{\#[Windy = no|Play = yes]}{\#[Play = yes]} = \frac{6}{9} = .66$$

$$P[yes|\omega_2] = \frac{\#[Windy = yes|Play = no]}{\#[Play = no]} = \frac{3}{5} = .6$$
$$P[no|\omega_2] = \frac{\#[Windy = no|Play = no]}{\#[Play = no]} = \frac{2}{5} = .4$$

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

Question: Will someone be playing on a sunny, mild, highly humid, and windy day?
This has never happened before, so how should we know???

Outlook	Temp	Humidity	Windy	Play?
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Example: Tennis dataset

Question: Will someone be playing on a sunny, mild, highly humid, and windy day?
This has never happened before, so how should we know???
We could look at the likelihood-ratio:

$$\frac{P[sunny|\omega_1] \times P[mild|\omega_1] \times P[high|\omega_1] \times P[yes|\omega_1]}{P[sunny|\omega_2] \times P[mild|\omega_2] \times P[high|\omega_2] \times P[yes|\omega_2]} > \frac{P[\omega_2]}{P[\omega_1]}$$

We calculated all the necessary numbers:

$P[sunny\|\omega_1]$	$P[overcast\|\omega_1]$	$P[rainy\|\omega_1]$	$P[sunny\|\omega_2]$	$P[overcast\|\omega_2]$	$P[rainy\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$
$P[hot\|\omega_1]$	$P[mild\|\omega_1]$	$P[cool\|\omega_1]$	$P[hot\|\omega_2]$	$P[mild\|\omega_2]$	$P[cool\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
$P[high\|\omega_1]$	$P[normal\|\omega_1]$	$P[high\|\omega_2]$	$P[normal\|\omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$		
$P[yes\|\omega_1]$	$P[no\|\omega_1]$	$P[yes\|\omega_2]$	$P[no\|\omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$		

$P[\omega_1]$	$P[\omega_2]$
$\frac{9}{14}$	$\frac{5}{14}$

Example: Tennis dataset

Question: Will someone be playing on a sunny, mild, highly humid, and windy day?
This has never happened before, so how should we know???
We could look at the likelihood-ratio:

$$\frac{\frac{2}{9} \times \frac{4}{9} \times \frac{3}{9} \times \frac{3}{9}}{\frac{3}{5} \times \frac{2}{5} \times \frac{4}{5} \times \frac{3}{5}} > \frac{\frac{5}{14}}{\frac{9}{14}}$$

$P[sunny\|\omega_1]$	$P[overcast\|\omega_1]$	$P[rainy\|\omega_1]$	$P[sunny\|\omega_2]$	$P[overcast\|\omega_2]$	$P[rainy\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$
$P[hot\|\omega_1]$	$P[mild\|\omega_1]$	$P[cool\|\omega_1]$	$P[hot\|\omega_2]$	$P[mild\|\omega_2]$	$P[cool\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$P[high\|\omega_1]$	$P[normal\|\omega_1]$	$P[high\|\omega_2]$	$P[normal\|\omega_2]$
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$
$P[yes\|\omega_1]$	$P[no\|\omega_1]$	$P[yes\|\omega_2]$	$P[no\|\omega_2]$
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$

$P[\omega_1]$	$P[\omega_2]$
$\frac{9}{14}$	$\frac{5}{14}$

Example: Tennis dataset

Question: Will someone be playing on a sunny, mild, highly humid, and windy day?
This has never happened before, so how should we know???
We could look at the likelihood-ratio:

$$\frac{45000}{472392} > \frac{70}{126}$$

$P[sunny\|\omega_1]$	$P[overcast\|\omega_1]$	$P[rainy\|\omega_1]$	$P[sunny\|\omega_2]$	$P[overcast\|\omega_2]$	$P[rainy\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$

$P[hot\|\omega_1]$	$P[mild\|\omega_1]$	$P[cool\|\omega_1]$	$P[hot\|\omega_2]$	$P[mild\|\omega_2]$	$P[cool\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$P[high\|\omega_1]$	$P[normal\|\omega_1]$	$P[high\|\omega_2]$	$P[normal\|\omega_2]$
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$

$P[yes\|\omega_1]$	$P[no\|\omega_1]$	$P[yes\|\omega_2]$	$P[no\|\omega_2]$
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$

$P[\omega_1]$	$P[\omega_2]$
$\frac{9}{14}$	$\frac{5}{14}$

Example: Tennis dataset

Question: Will someone be playing on a sunny, mild, highly humid, and windy day?
This has never happened before, so how should we know???
We could look at the likelihood-ratio:

$$.095 > .55$$

So the answer to the question is no.

$P[sunny\|\omega_1]$	$P[overcast\|\omega_1]$	$P[rainy\|\omega_1]$	$P[sunny\|\omega_2]$	$P[overcast\|\omega_2]$	$P[rainy\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$
$P[hot\|\omega_1]$	$P[mild\|\omega_1]$	$P[cool\|\omega_1]$	$P[hot\|\omega_2]$	$P[mild\|\omega_2]$	$P[cool\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
$P[high\|\omega_1]$	$P[normal\|\omega_1]$	$P[high\|\omega_2]$	$P[normal\|\omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$		
$P[yes\|\omega_1]$	$P[no\|\omega_1]$	$P[yes\|\omega_2]$	$P[no\|\omega_2]$	$P[\omega_1]$	$P[\omega_2]$
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{9}{14}$	$\frac{5}{14}$

Example: Tennis dataset

Now let's look at the same question for an overcast day

$$\frac{\frac{4}{9} \times \frac{4}{9} \times \frac{3}{9} \quad \times \frac{3}{9}}{\frac{0}{5} \quad \times \frac{2}{5} \quad \times \frac{4}{5} \times \frac{3}{5}} > \frac{\frac{5}{14}}{\frac{9}{14}}$$

$P[sunny\|\omega_1]$	$P[overcast\|\omega_1]$	$P[rainy\|\omega_1]$	$P[sunny\|\omega_2]$	$P[overcast\|\omega_2]$	$P[rainy\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$
$P[hot\|\omega_1]$	$P[mild\|\omega_1]$	$P[cool\|\omega_1]$	$P[hot\|\omega_2]$	$P[mild\|\omega_2]$	$P[cool\|\omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
$P[high\|\omega_1]$	$P[normal\|\omega_1]$	$P[high\|\omega_2]$	$P[normal\|\omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$		
$P[yes\|\omega_1]$	$P[no\|\omega_1]$	$P[yes\|\omega_2]$	$P[no\|\omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$		

$P[\omega_1]$	$P[\omega_2]$
$\frac{9}{14}$	$\frac{5}{14}$

Example: Tennis dataset

Now let's look at the same question for an overcast day

$$\frac{45000}{0} > \frac{70}{126}$$

This leads to a division by zero! So, what happened here?
We never observed no one playing on an overcast day.

$P[sunny \mid \omega_1]$	$P[overcast \mid \omega_1]$	$P[rainy \mid \omega_1]$	$P[sunny \mid \omega_2]$	$P[overcast \mid \omega_2]$	$P[rainy \mid \omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{3}{5}$	$\frac{0}{5}$	$\frac{2}{5}$
$P[hot \mid \omega_1]$	$P[mild \mid \omega_1]$	$P[cool \mid \omega_1]$	$P[hot \mid \omega_2]$	$P[mild \mid \omega_2]$	$P[cool \mid \omega_2]$
$\frac{2}{9}$	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$
$P[high \mid \omega_1]$	$P[normal \mid \omega_1]$	$P[high \mid \omega_2]$	$P[normal \mid \omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{4}{5}$	$\frac{1}{5}$		
$P[yes \mid \omega_1]$	$P[no \mid \omega_1]$	$P[yes \mid \omega_2]$	$P[no \mid \omega_2]$		
$\frac{3}{9}$	$\frac{6}{9}$	$\frac{3}{5}$	$\frac{2}{5}$		

$P[\omega_1]$	$P[\omega_2]$
$\frac{9}{14}$	$\frac{5}{14}$

Laplace smoothing

- To determine a discrete probability from counting samples, we can proceed as follows:
  - We count $N_1, \ldots, N_m$ samples for each category
  - Typically, we would then determine the probability of each category as the fraction of the number of samples divided by the total number of samples, i.e.

$$P[i] = \frac{N_i}{\sum_{i=1}^{m} N_i}$$

- This can lead to probabilities being zero in case we did not count any samples for this category, i.e. $N_i = 0 \Rightarrow P[i] = 0$

- While not having counted any sample for a category can happen, zero probability means that it is impossible to ever count any sample
- This is a strong statement we usually want to avoid

- For that reason we can artificially add a fixed number α of samples to each category
- The resulting probabilities are then always non-zero

$$P[i] = \frac{N_i + \alpha}{\sum_{i=1}^{m}(N_i + \alpha)} = \frac{N_i + \alpha}{\sum_{i=1}^{m} N_i + m\alpha}$$

- This procedure leads to a smoothing of the probabilities, with extremely large α having the effect of all probabilities to be almost equal

Example: Tennis dataset

Let's go back to the computation of $P[x_1|\omega_2]$

The number of samples were

$$\#[Outlook = sunny|Play = no] = 3$$
$$\#[Outlook = overcast|Play = no] = 0$$
$$\#[Outlook = rainy|Play = no] = 2$$

Therefore the likelihood $P[x_1|\omega_2]$ was

$$P[sunny|\omega_2] = \frac{\#[Outlook = sunny|Play = no]}{\#[Play = no]} = \frac{3}{5} = .6$$
$$P[overcast|\omega_2] = \frac{\#[Outlook = overcast|Play = no]}{\#[Play = no]} = \frac{0}{5} = 0$$
$$P[rainy|\omega_2] = \frac{\#[Outlook = rainy|Play = no]}{\#[Play = no]} = \frac{2}{5} = .4$$

We now apply Laplace smoothing and add 1 sample for each term

$$P[sunny|\omega_2] = \frac{\#[Outlook = sunny|Play = no] + 1}{\#[Play = no] + 3} = \frac{4}{8} = .5$$
$$P[overcast|\omega_2] = \frac{\#[Outlook = overcast|Play = no] + 1}{\#[Play = no] + 3} = \frac{1}{8} = .125$$
$$P[rainy|\omega_2] = \frac{\#[Outlook = rainy|Play = no] + 1}{\#[Play = no] + 3} = \frac{3}{8} = .375$$

The issue with the 0 disappeared, because it drew some probabilities from its neighbours.

Outlook	**Temp**	**Humidity**	**Windy**	**Play?**
sunny	hot	high	no	no
sunny	hot	high	yes	no
overcast	hot	high	no	yes
rainy	mild	high	no	yes
rainy	cool	normal	no	yes
rainy	cool	normal	yes	no
overcast	cool	normal	yes	yes
sunny	mild	high	no	no
sunny	cool	normal	no	yes
rainy	mild	normal	no	yes
sunny	mild	normal	yes	yes
overcast	mild	high	yes	yes
overcast	hot	normal	no	yes
rainy	mild	high	yes	no

Numerical stability of naïve Bayesian classification

- While the smoothing took care of the zeros, we still need to calculate products over sometimes very small (non-zero) numbers

$$\prod_{i=1}^{n} P\left[x_i \middle| \omega_j\right]$$

- Because of numerical precision limitations, these products can evaluate to zero, even if all factors are greater than 0

- To avoid this complication, always instead sum up log-likelihoods by making use of the following equality

$$\prod_{i=1}^{n} P\left[x_i \middle| \omega_j\right] = \exp\left[\log\left[\prod_{i=1}^{n} P\left[x_i \middle| \omega_j\right]\right]\right] = \exp\left[\sum_{i=1}^{n} \log\left[P\left[x_i \middle| \omega_j\right]\right]\right]$$

Example: Using log-likelihoods in naïve Bayesian classification

$$\prod_{i=1}^{n} P[x_i|\omega_j] = \exp\left[\log\left[\prod_{i=1}^{n} P[x_i|\omega_j]\right]\right] = \exp\left[\sum_{i=1}^{n} \log\left[P[x_i|\omega_j]\right]\right]$$

- In the previous example we had to calculate

$$\frac{4}{9} \times \frac{4}{9} \times \frac{3}{9} \times \frac{3}{9} = .44 \times .44 \times .33 \times .33 = .022$$

- If any of the factors becomes too small, this product will evaluate to zero and not consider the other factors at all

- Using logarithms instead, we could have also computed the product as

$$\exp[\log[.44] + \log[.44] + \log[.33] + \log[.33]] = \exp[-.82 - .82 - 1.1 - 1.1] = .022$$

- This is much more robust to small numbers and should always be calculated like this

Thank you for your attention