DSC 40A

Theoretical Foundations of Data Science I

Announcements

- Homework 6 due today
- Homework 5 grades released
- Homework 7 will be released Wednesday 11/27 and due 12/6.

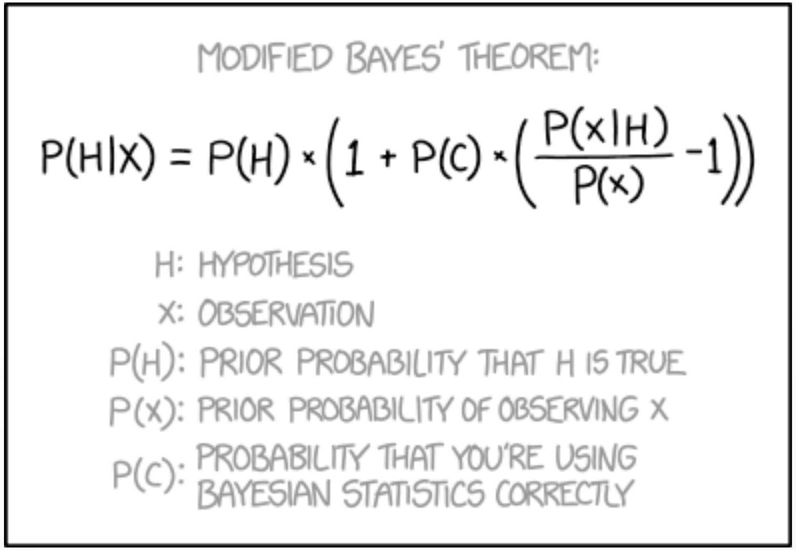
Question

Answer at q.dsc40a.com

Remember, you can always ask questions at q.dsc40a.com!
If the direct link doesn't work, click the "Lecture Questions" link in the top right corner of dsc40a.com.

Agenda

- Bayes Theorem
- Naïve Bayes Classifier



Source: xkcd

Bayes Theorem

Last Week

- We defined Bayes' Theorem:

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

- Bayes' Theorem describes how to update the probability of one event given that another has occurred.

Bayes' Theorem

Bayes' Theorem follows from the multiplication rule, or conditional probability.

$$P(A) * P(B|A) = P(A \text{ and } B) = P(B) * P(A|B)$$

Bayes' Theorem:

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

law of total probability

$$= \frac{P(A|B) * P(B)}{P(B) * P(A|B) + P(\overline{B}) * P(A|\overline{B})}$$

not B

Bayes' Theorem

For hypothesis H and evidence (data) E

$$P(H \mid E) = \frac{P(E|H)\,P(H)}{P(E)} = \frac{P(E|H) \cdot P(H)}{P(E|H)\,P(H) + P(E|\bar{H}) \cdot P(\bar{H})}$$

- $P(H)$ - prior, initial probability before E is observed
- $P(H|E)$ - posterior, probability of H after E is observed
- $P(E|H)$ - likelihood, probability of E if the hypothesis is true
- $P(E)$ - marginal, probability of E regardless of H

$$P(E|\bar{H}) \neq P(\bar{E}|H)$$

The likelihood function is a function of E, while the posterior probability is a function of H.

Bayes' Theorem: Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\sim H)P(\sim H)}$$

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

What is your first guess?

A. Close to 95%
70% B. Close to 85%
C. Close to 40%
D. Close to 15%

Bayes' Theorem: Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\sim H)P(\sim H)}$$

$\bar{H}$, not H

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Now, calculate it and choose the best answer.

A. Close to 95%
B. Close to 85%
C. Close to 40%
D. Close to 15%

E - positive drug test

H - uses steroids

$\bar{H}$ - doesn't use steroids

Bayes' Theorem: Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\sim H)P(\sim H)} = \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.15 \cdot 0.9} \approx 0.41$$

A manufacturer claims that its drug test will **detect steroid use 95% of the time**. What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Solution:

$P(E|H) = 95\%$ (TP)

H: used steroids

$P(E|\bar{H}) = 15\%$ (FP)

$\Longrightarrow P(H|E) = 41\%$

E: tested positive

$P(H) = 10\%$ $\quad P(\bar{H}) = 90\%$

Bayes' Theorem: Example

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\sim H)P(\sim H)}$$

A manufacturer claims that its drug test will **detect steroid use 95% of the time.** What the company does not tell you is that 15% of all steroid-free individuals also test positive (the false positive rate). 10% of the Tour de France bike racers use steroids. Your favorite cyclist just tested positive. What's the probability that he used steroids?

Solution:

H: used steroids

E: tested positive

Despite manufacturer's claims, only **41% chance** that cyclist used steroids.

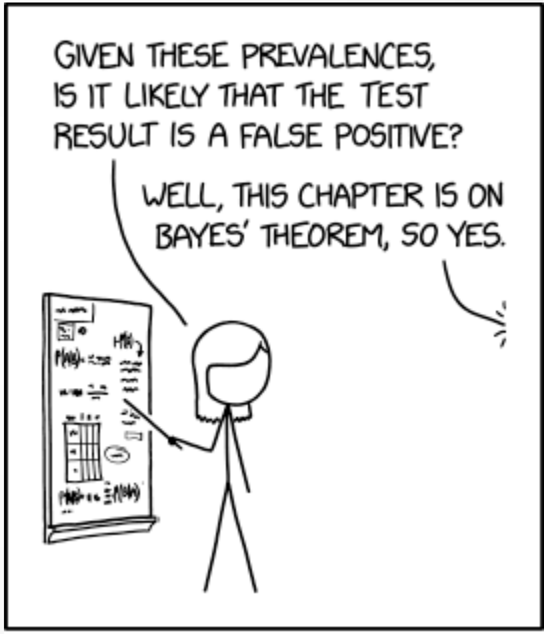
Bayes' Theorem: Example

Example

- 1% of people have a certain genetic defect
- 90% of tests accurately detect the gene (true positives).
- 7% of the tests are false positives.

H - has genetic disorder
E - positive test result

If Olaf gets a positive test result, what are the odds he actually has the genetic defect?



SOMETIMES, IF YOU UNDERSTAND BAYES' THEOREM WELL ENOUGH, YOU DON'T NEED IT.

Bayes' Theorem: Example

- Hypothesis: Olaf has the gene, $P(H) = 0.01$
- Evidence: Olaf got a positive test result, $P(E)$
- True positive: Probability of positive test result if someone has the gene $P(E|H) = 0.9$
- False positive: Probability of positive test result if someone doesn't have the gene $P(E|\bar{H}) = 0.07$

Bayes' Theorem: Example

Calculate

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)\,P(H)}{P(E|H)\cdot P(H) + P(E|\bar{H})\cdot P(\bar{H})}$$

$$= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.07 \cdot 0.99} = \ldots = 0.115$$

$1 - 0.01$

The probability that Olaf has the gene is only 11.5% despite the positive test result!

Bayes' Theorem: Example

What happens if there are less false positives?
Consider $P(E|\overline{H}) = 0.02$:

$$P(H|E) = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + \boxed{0.02 \cdot 0.99}} = \ldots \approx 31\%$$

The probability that Olaf has the gene is now ___31%___.

Bayes' Theorem: Example

What happens if there are more true positives?
Consider $P(E|H) = 0.95$:

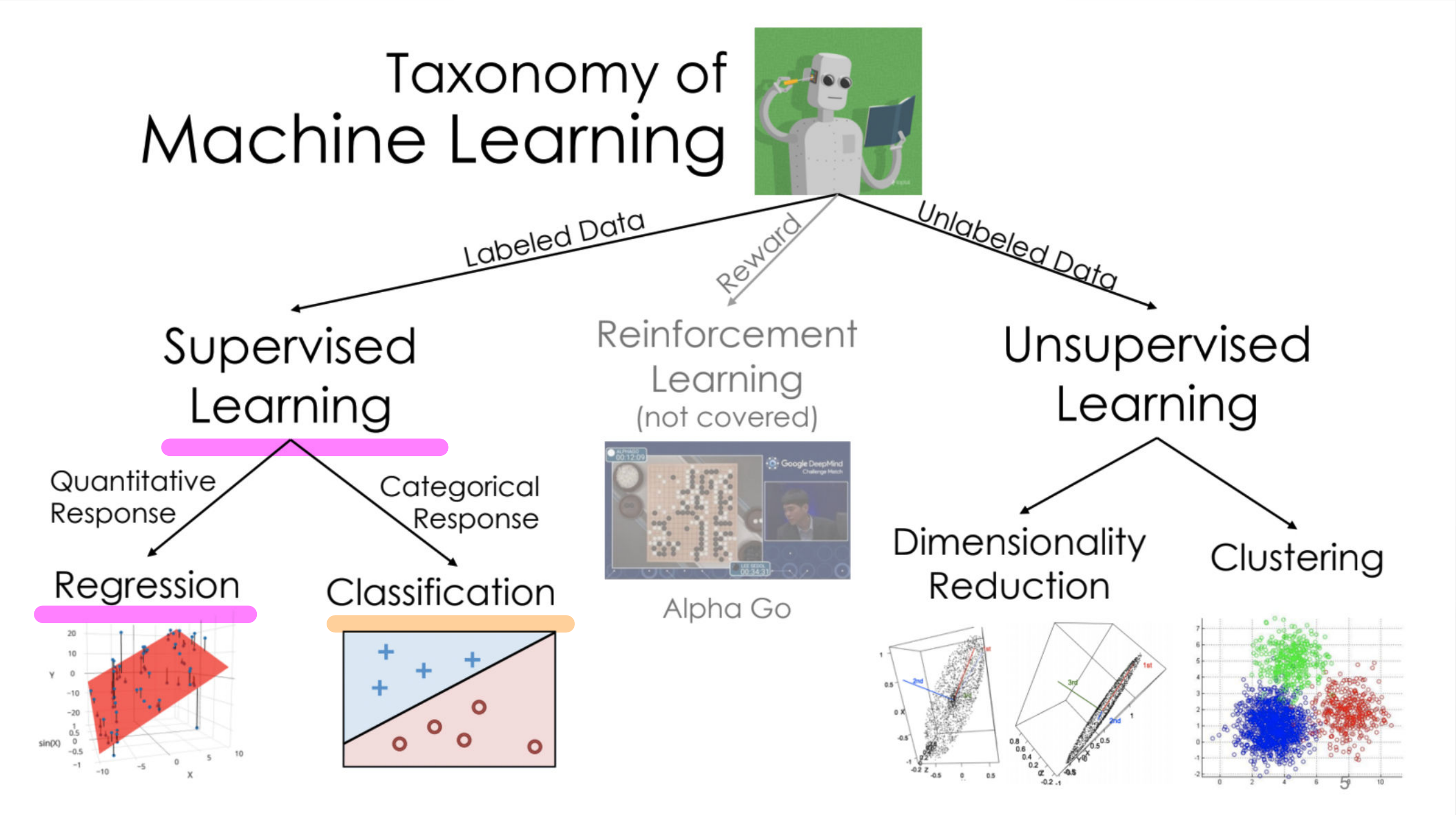
$$P(H|E) = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.11 + 0.07 \cdot 0.99} = \ldots = 0.12$$

Improving the accuracy of true positives raised the probability that Olaf has the gene to 12%.

Naïve Bayes Classifier

Today

- Using Bayes' Theorem to solve the classification problem



Preview: Bayes' Theorem for Classification

Bayes’ Theorem is very useful for classification problems, where we want to predict a class based on some features.

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

B = belonging to a certain class
A = having certain features

$$P(\text{class}|\text{features}) = \frac{P(\text{features}|\text{class}) * P(\text{class})}{P(\text{features})}$$

Classification

- Making predictions based on examples (training data)
- Response variable is categorical
- Categories are called *classes*
- Examples:
    - decide whether patient has kidney disease
    - identify handwritten digits MNIST $\{0,1,2,3,4,5,6,7,8,9\}$
    - determine whether an avocado is ripe
    - predict whether credit card activity is fraudulent

features
class

Example

Color	Ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

You have a green-black avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Which class would you predict?

A. ripe
B. unripe

Example

Color	Ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

You have a green-black avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Strategy: Calculate two probabilities:

$$P(\text{ripe} \mid \text{green-black}) \approx \frac{\#(\text{ripe and green black})}{\#(\text{green black})} = \frac{3}{5}$$

$$P(\text{unripe} \mid \text{green-black}) \approx \frac{2}{5}$$

Then choose the class according to the **larger** of these two probabilities.

$$\frac{3}{5} > \frac{2}{5} \Rightarrow \text{ripe}$$

Bayes' Theorem for Classification

Bayes' Theorem gives another strategy for predicting the class given features.

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

ain class

ures

B – belonging to certain class

A – features

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{\boxed{P(\text{features})}}$$

Bayes' Theorem for Classification

Bayes’ Theorem gives another strategy for predicting the class given features.

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

ain class

ures

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{\boxed{P(\text{features})}}$$

Can all be estimated from the training data

Avocado Ripeness

Color	Ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

You have a green-black avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{\boxed{P(\text{features})}}$$

$$P(\text{green black} | \text{ripe}) = \frac{3}{7}$$

$$P(\text{ripe}) = \frac{7}{11}$$

← ripe avocados

← total avocados

$$P(\text{green black}) = \frac{5}{11}$$

$$= \frac{\frac{3}{\cancel{7}} \cdot \frac{\cancel{7}}{\cancel{11}}}{\frac{5}{\cancel{11}}} = \frac{3}{5}$$

$$P(\text{unripe} | \text{green black}) = \ldots = \frac{2}{5}$$

Avocado Ripeness

Color	Ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

You have a green-black avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{\boxed{P(\text{features})}}$$

$$P(\overline{\text{class}}\,|\,\text{features}) = \frac{\qquad}{P(\text{features})}$$

Avocado Ripeness

Color	Ripeness
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	unripe
purple-black	ripe
bright green	unripe
green-black	ripe
purple-black	ripe
green-black	ripe
green-black	unripe
purple-black	ripe

You have a green-black avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{P(\text{features})}$$

Shortcut: Both probabilities have same denominator. To find larger one, choose one with larger numerator.

P(ripe | green-black) $\underbrace{\propto}_{\text{proportional}} \frac{3}{7} \cdot \frac{7}{11} = \frac{3}{11}$ ✓

P(unripe | green-black) $\propto \frac{2}{4} \cdot \frac{4}{11} = \frac{2}{11}$

3 features

More Features

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Strategy: Calculate two probabilities:

P(ripe | firm, green-black, Zutano)

Problem: no firm & green black & Zutano in dataset

P(unripe | firm, green-black, Zutano)

Then choose the class according to the **larger** of these two probabilities.

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Problem: We have not seen an avocado with all these features. Both probabilities will be undefined.

P(ripe | firm, ~~green-black~~, Zutano)

P(unripe | firm, ~~green-black~~, Zutano)

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{P(\text{features})}$$

Solution: Use Bayes' Theorem, plus a simplifying assumption, to calculate the two numerators.

conditional independence

⇒ Naive Bayes Classifier!

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * P(\text{class})}{P(\text{features})}$$

Simplifying assumption: Within a given class, the features are independent.

P(firm, green-black, Zutano | ripe) =
P(firm | ripe)*P(green-black | ripe)*P(Zutano | ripe)

Conditional Independence

- Recall that A and B are independent if

$$P(A \text{ and } B) = P(A) * P(B)$$

- A and B are conditionally independent given C if

$$P((A \text{ and } B)|C) = P(A|C) * P(B|C)$$

- Given that C occurs, this says that A and B are independent of one another.

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{P(\text{features})}$$

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

Assuming conditional independence of features given the class, calculate P(firm, green-black, Zutano | unripe).

A. 0
B. 1/4
C. 3/16
D. 1 - (1/7*3/7*2/7)

Avocado Ripeness

Color	Softness	Variety	Ripeness
bright green	firm	Zutano	unripe
green-black	medium	Hass	ripe
purple-black	firm	Hass	ripe
green-black	medium	Hass	unripe
purple-black	soft	Hass	ripe
bright green	firm	Zutano	unripe
green-black	soft	Zutano	ripe
purple-black	soft	Hass	ripe
green-black	soft	Zutano	ripe
green-black	firm	Hass	unripe
purple-black	medium	Hass	ripe

You have a firm green-black Zutano avocado. Based on this data, would you predict that your avocado is ripe or unripe?

$$P(\text{class}|\text{features}) = \frac{\boxed{P(\text{features}|\text{class})} * \boxed{P(\text{class})}}{P(\text{features})}$$

Naïve Bayes Algorithm

- Bayes' Theorem shows how to calculate P(class | features).

$$P(\text{class}|\text{features}) = \frac{P(\text{features}|\text{class}) * P(\text{class})}{P(\text{features})}$$

- Rewrite the numerator, using the naïve assumption of conditional independence of features given the class.
- Estimate each term in the numerator based on the training data.
- Select class based on whichever has the larger numerator.

Summary

- The Naïve Bayes algorithm gives a strategy for classifying data according to its features.
- It relies on an assumption of conditional independence of the features.
- **Next time:** application to text classification