Lecture 16 – Naive Bayes

DSC 40A, Fall 2021 @ UC San Diego

Suraj Rampure, with help from **many others**

Announcements

- Fill out Survey 7 if you haven't already!
- **No groupwork or discussion section this week.**
- Homework 8, the final homework, will come out by Thursday and is due **Friday 12/3 at 11:59pm**.
- No office hours on Wednesday, Thursday, Friday, or Saturday this week.
- Lots of office hours during the last week of class — **start studying for the Final Exam early!**
    - Wednesday, 12/8, 11:30AM-2:30PM, remote (same format as midterm).

Agenda

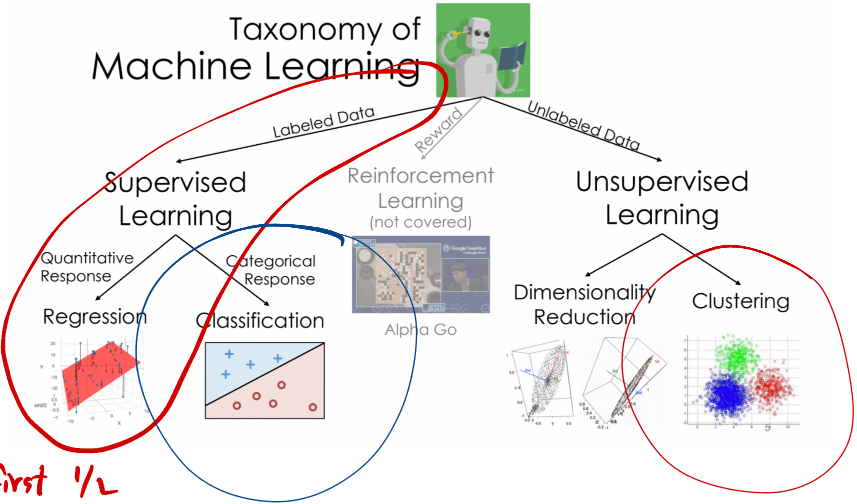
- Classification.
- Classification and conditional independence.
- Naive Bayes.

Recap: Bayes' theorem, independence, and conditional independence

- Bayes' theorem: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
- A and B are **independent** if $P(A \cap B) = P(A) \cdot P(B)$.
- A and B are **conditionally independent** given C if $P((A \cap B)|C) = P(A|C) \cdot P(B|C)$.
  - In general, there is no relationship between independence and conditional independence.
  - **See the Campuswire post on conditional independence if you're still shaky on the concept.**

Classification

Taxonomy of machine learning



first 1/2 of the class

today

k-Means Clustering

Classification problems

- Like with regression, we're interested in ~~mkaing~~ making predictions based on data we've already collected (called **training data**).

- The difference is that the response variable is **categorical**.

- Categories are called **classes**.

- Example classification problems:
  - Deciding whether a patient has kidney disease.
  - Identifying handwritten digits.
  - Determining whether an avocado is ripe.
  - Predicting whether credit card activity is fraudulent.

Example: avocados

You have a green-black avocado, and want to know if it is ripe.

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

Question: Based on this data, would you predict that your avocado is ripe or unripe?

$\frac{3}{5}$ are ripe

$\frac{2}{5}$ are unripe

$\rightarrow$ $\boxed{\text{predict ripe}}$

Example: avocados

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

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

Strategy: Calculate two probabilities:

$$P(\text{ripe} | \text{green-black})$$

$$P(\text{unripe} | \text{green-black})$$

Then, predict the class with a **larger** probability.

Estimating probabilities

- We would like to determine $P(\text{ripe}|\text{green-black})$ and $P(\text{unripe}|\text{green-black})$ for all avocados in the universe.

- All we have is a single dataset, which is a **sample** of all avocados in the universe.

- We can estimate these probabilities by using sample proportions.

$$\underbrace{P(\text{ripe}|\text{green-black})}_{\text{population parameter}} \approx \frac{\text{\# ripe green-black avocados in sample}}{\text{\# green-black avocados in sample}}$$

- Per the **law of large numbers** in DSC 10, larger samples lead to more reliable estimates of population parameters.

Example: avocados

$$\frac{P(\text{ripe and green-black})}{P(\text{green-black})}$$

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

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

$P(\text{ripe}|\text{green-black})$ =

$$\frac{\#\text{ ripe green-black}}{\#\text{ green-black}} = \frac{3}{5}$$

$P(\text{unripe}|\text{green-black})$ =

$$\frac{\#\text{ unripe green-black}}{\#\text{ green-black}} = \frac{2}{5}$$

Bayes' theorem for classification

- Suppose that A is the event that an avocado has certain features, and B is the event that an avocado belongs to a certain class. Then, by Bayes' theorem:

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

ripe green-black

- More generally:

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

- What's the point?
  - Usually, it's not possible to estimate $P(\text{class}|\text{features})$ directly from the data we have.
  - Instead, we have to estimate $P(\text{class})$, $P(\text{features}|\text{class})$, and $P(\text{features})$ separately.

Example: avocados

gb = green-black

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

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

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

$$P(ripe|gb) = \frac{P(ripe) \cdot P(gb|ripe)}{P(gb)}$$

① $P(ripe) = \frac{7}{11}$

② $P(gb|ripe) = \frac{3}{7}$

③ $P(gb) = \frac{5}{11}$

$$P(ripe|gb) = \frac{\frac{\cancel{7}}{11} \cdot \frac{3}{\cancel{7}}}{\frac{5}{11}} = \frac{3}{5}$$

Example: avocados

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

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

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

$$P(\text{unripe}|gb) = \frac{P(\text{unripe})\,P(gb|\text{unripe})}{P(gb)}$$

$$= \frac{\frac{4}{11}\cdot\frac{2}{4}}{\frac{5}{11}} = \frac{2}{5}$$

Example: avocados

You have a green-black avocado, and want to know if it is ripe. Based on this data, would you predict that your avocado is ripe or unripe?

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

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

Shortcut: Both probabilities have the same denominator. The larger one is the one with the larger numerator.

proportional to

$P(\text{ripe}|\text{green-black})$

$$\propto \frac{7}{11} \cdot \frac{3}{7} = \frac{3}{11}$$

$P(\text{unripe}|\text{green-black})$

$$\propto \frac{4}{11} \cdot \frac{2}{4} = \frac{2}{11}$$

Classification and conditional independence

Example: avocados, but with 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?

3 features

Example: avocados, but with 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?

Strategy: Calculate $P(\text{ripe}|\text{features})$ and $P(\text{unripe}|\text{features})$ and choose the class with the **larger** probability.

$$P(\text{ripe}|\text{firm, green-black, Zutano})$$

$$P(\text{unripe}|\text{firm, green-black, Zutano})$$

Example: avocados, but with 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?

Issue: We have not seen a firm green-black Zutano avocado before.

$$\frac{\text{\# ripe firm gb Zutano}}{\text{\# firm gb Zutano}} = \frac{0}{0}$$

This means that $P(\text{ripe}|\text{firm, green-black, Zutano})$ and $P(\text{unripe}|\text{firm, green-black, Zutano})$ are undefined.

A simplifying assumption

comma means "and"

- We want to find $P(\text{ripe}|\text{firm, green-black, Zutano})$, but there are no firm green-black Zutano avocados in our dataset.

- Bayes' theorem tells us this probability is equal to

P(class) · P(features | class)

$$P(\text{ripe}|\text{firm, green-black, Zutano}) = \frac{P(\text{ripe}) \cdot P(\text{firm, green-black, Zutano}|\text{ripe})}{P(\text{firm, green-black, Zutano})}$$

P(class | features)

P(features)

- **Key idea:** **Assume** that features are **conditionally independent** given a class (e.g. ripe).

$$P(\text{firm, green-black, Zutano}|\text{ripe}) = P(\text{firm}|\text{ripe}) \cdot P(\text{green-black}|\text{ripe}) \cdot P(\text{Zutano}|\text{ripe})$$

conditional independence:

$$\underbrace{P((A \cap B)}_{\text{features}} \mid \underbrace{C}_{\text{class}}) = P(A \mid C)\, P(B \mid C)$$

$P(\text{firm, green-black, Zutano} \mid \text{ripe})$

$= P(\text{firm} \mid \text{ripe}) \cdot$

$P(\text{green-black} \mid \text{ripe}) \cdot$

$P(\text{Zutano} \mid \text{ripe})$

Example: avocados, but with 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?

$$P(\text{ripe}\,|\,\text{firm, green-black, Zutano}) = \frac{P(\text{ripe}) \cdot P(\text{firm, green-black, Zutano}\,|\,\text{ripe})}{P(\text{firm, green-black, Zutano})}$$

$$\propto P(\text{ripe}) \cdot P(\text{firm, green-black, Zutano}\,|\,\text{ripe})$$

$$= P(\text{ripe}) \cdot P(\text{firm}\,|\,\text{ripe}) \cdot P(\text{green-black}\,|\,\text{ripe}) \cdot P(\text{Zutano}\,|\,\text{ripe})$$

$$= \frac{\cancel{7}}{11} \cdot \frac{1}{\cancel{7}} \cdot \frac{3}{7} \cdot \frac{2}{7} = \frac{6}{11 \cdot 49} = \frac{6}{539}$$

Example: avocados, but with 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?

$$P(\text{unripe}|\text{firm, green-black, Zutano}) = \frac{P(\text{unripe}) \cdot P(\text{firm, green-black, Zutano}|\text{unripe})}{P(\text{firm, green-black, Zutano})}$$

$$\propto P(\text{unripe}) \cdot P(\text{firm}|\text{unripe}) \cdot P(\text{green-black}|\text{unripe}) \cdot P(\text{Zutano}|\text{unripe})$$

$$= \frac{4}{11} \cdot \frac{3}{4} \cdot \frac{\cancel{2}}{\cancel{4}} \cdot \frac{\cancel{2}}{\cancel{4}} = \frac{3}{44} = \frac{6}{88}$$

Conclusion

- The numerator of $P(\text{ripe}|\text{firm, green-black, Zutano})$ is $\frac{6}{539}$.
- The numerator of $P(\text{unripe}|\text{firm, green-black, Zutano})$ is $\frac{6}{88}$.
  - Both probabilities have the same denominator, $P(\text{firm, green-black, Zutano})$.
  - Since we're just interested in seeing which one is larger, we can ignore the denominator and compare numerators.
- Since the numerator for unripe is **larger** than the numerator for ripe, we **predict that our avocado is unripe**.

Naive Bayes

Naive Bayes classifier

- We want to predict a class, given certain features.
- Using Bayes' theorem, we write

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

- For each class, we compute the numerator using the **naive assumption of conditional independence of features given the class**.
- We estimate each term in the numerator based on the training data.
- We predict the class with the largest numerator.
  - Works if we have multiple classes, too!

na·ive

/nīˈēv/

adjective

(of a person or action) showing a lack of experience, wisdom, or judgment.
"the rather naive young man had been totally misled"

- (of a person) natural and unaffected; innocent.
  "Andy had a sweet, naive look when he smiled"
  Similar: innocent · unsophisticated · artless · ingenuous · inexperienced
- of or denoting art produced in a straightforward style that deliberately rejects sophisticated artistic techniques and has a bold directness resembling a child's work, typically in bright colors with little or no perspective.

Example: comic characters

ALIGN	SEX	COMPANY
Bad	Male	Marvel
Neutral	Male	Marvel
Good	Male	Marvel
Bad	Male	DC
Good	Female	Marvel
Bad	Male	DC
Good	Male	DC
Bad	Male	Marvel
Good	Female	Marvel
Bad	Female	Marvel

My favorite character is a male Marvel character. Using Naive Bayes, would we predict that my favorite character is bad, good, or neutral?

ALIGN	SEX	COMPANY
Bad	Male	Marvel
Neutral	Male	Marvel
Good	Male	Marvel
Bad	Male	DC
Good	Female	Marvel
Bad	Male	DC
Good	Male	DC
Bad	Male	Marvel
Good	Female	Marvel
Bad	Female	Marvel

male Marvel character
bad? good? neutral?

$$P(\text{bad} \mid \text{male}, \text{Marvel})$$

$$\propto P(\text{bad}) \cdot P(\text{male}, \text{Marvel} \mid \text{bad})$$

$$\overset{\text{assp.}}{=} P(\text{bad}) \cdot P(\text{male} \mid \text{bad}) \cdot P(\text{Marvel} \mid \text{bad})$$

$$= \frac{\cancel{5}}{\cancel{10}_{5}} \cdot \frac{\cancel{4}^{2}}{\cancel{5}_{5}} \cdot \frac{3}{5} = \frac{6}{25}$$

$$P(\text{good} \mid \text{male}, \text{Marvel}) \propto P(\text{good}) \cdot P(\text{male} \mid \text{good}) \cdot P(\text{Marvel} \mid \text{good})$$

$$= \frac{\cancel{4}}{\cancel{10}_{5}} \cdot \frac{\cancel{2}^{1}}{\cancel{4}} \cdot \frac{3}{4} = \frac{3}{20} = \frac{6}{40}$$

$$P(\text{neutral} \mid \text{male}, \text{Marvel}) \propto P(\text{neutral}) \cdot P(\text{male} \mid \text{neutral}) \cdot P(\text{Marvel} \mid \text{neutral})$$

$$= \frac{1}{10} \cdot 1 \cdot 1 = \frac{1}{10} = \frac{6}{60}$$

$\Rightarrow$ predict character is bad!!!

Example: comic characters

ALIGN	SEX	COMPANY
Bad	Male	Marvel
Neutral	Male	Marvel
Good	Male	Marvel
Bad	Male	DC
Good	Female	Marvel
Bad	Male	DC
Good	Male	DC
Bad	Male	Marvel
Good	Female	Marvel
Bad	Female	Marvel

My other favorite character is a female Marvel character. Using Naive Bayes, would we predict that my favorite character is bad, good, or neutral?

other

ALIGN	SEX	COMPANY
Bad	Male	Marvel
Neutral	Male	Marvel
Good	Male	Marvel
Bad	Male	DC
Good	Female	Marvel
Bad	Male	DC
Good	Male	DC
Bad	Male	Marvel
Good	Female	Marvel
Bad	Female	Marvel

$$\frac{0}{1} = 0$$

$$P(\text{neutral} \mid \text{female}, \text{Marvel}) \propto P(\text{neutral}) \cdot P(\text{female} \mid \text{neutral}) \cdot P(\text{Marvel} \mid \text{neutral})$$

Uh oh...

- There are no neutral female characters in the data set.

- The estimate $P(\text{female}|\text{neutral}) \approx \frac{\text{\# female neutral characters}}{\text{\# neutral characters}}$ is 0.

- The estimated numerator,
  $P(\text{neutral}) \cdot P(\text{female, Marvel}|\text{neutral}) =$
  $P(\text{neutral}) \cdot P(\text{female}|\text{neutral}) \cdot P(\text{Marvel}|\text{neutral})$,
  is also 0.

- But just because there isn't a neutral female character in the data set, doesn't mean they don't exist!

- **Idea:** Adjust the numerators and denominators of our estimate so that they're never 0.

Smoothing

Edit after lecture:
do <u>NOT</u> smooth any unconditional probabilities!

<u>ONLY</u> smooth conditional probabilities.

- **Without** smoothing:

$$P(\text{bad}) \approx \frac{\#\text{ bad}}{\#\text{ bad} + \#\text{ good} + \#\text{ neutral}}$$

$$P(\text{good}) \approx \frac{\#\text{ good}}{\#\text{ bad} + \#\text{ good} + \#\text{ neutral}}$$

$$P(\text{neutral}) \approx \frac{\#\text{ neutral}}{\#\text{ bad} + \#\text{ good} + \#\text{ neutral}}$$

- **With** smoothing:

$$P(\text{bad}) \approx \frac{\#\text{ bad} + 1}{\#\text{ bad} + 1 + \#\text{ good} + 1 + \#\text{ neutral} + 1}$$

$$P(\text{good}) \approx \frac{\#\text{ good} + 1}{\#\text{ bad} + 1 + \#\text{ good} + 1 + \#\text{ neutral} + 1}$$

$$P(\text{neutral}) \approx \frac{\#\text{ neutral} + 1}{\#\text{ bad} + 1 + \#\text{ good} + 1 + \#\text{ neutral} + 1}$$

Smoothing

only smooth conditional probabilities (like the ones on this slide)

- **Without** smoothing:

$$P(\text{female}|\text{neutral}) \approx \frac{\#\text{ female neutral}}{\#\text{ female neutral} + \#\text{ male neutral}}$$

$$P(\text{male}|\text{neutral}) \approx \frac{\#\text{ male neutral}}{\#\text{ female neutral} + \#\text{ male neutral}}$$

- **With** smoothing:

$$P(\text{female}|\text{neutral}) \approx \frac{\#\text{ female neutral} + 1}{\#\text{ female neutral} + 1 + \#\text{ male neutral} + 1}$$

$$P(\text{male}|\text{neutral}) \approx \frac{\#\text{ male neutral} + 1}{\#\text{ female neutral} + 1 + \#\text{ male neutral} + 1}$$

- When smoothing, we add 1 to the count of every group whenever we're estimating a probability.

Example: comic characters

not smoothed (only smooth conditionals)

Using smoothing, let's determine whether Naive Bayes would predict a female Marvel character to be bad, good, or neutral.

ALIGN	SEX	COMPANY
Bad	Male	Marvel
Neutral	Male	Marvel
Good	Male	Marvel
Bad	Male	DC
Good	Female	Marvel
Bad	Male	DC
Good	Male	DC
Bad	Male	Marvel
Good	Female	Marvel
Bad	Female	Marvel

$$P(bad \mid f, M) \propto P(bad) \cdot P(f \mid bad) \cdot P(M \mid bad)$$

$$= \left(\frac{5}{10}\right)\left(\frac{1+1}{1+1+4+1}\right)\left(\frac{3+1}{3+1+2+1}\right)$$

$$= \left(\frac{5}{10}\right) \cdot \left(\frac{2}{7}\right) \cdot \left(\frac{4}{7}\right) = \frac{4}{49}$$

$$P(good \mid f, M) \propto P(good) \cdot P(f \mid good) \cdot P(Marvel \mid good)$$

$$= \left(\frac{4}{10}\right)\left(\frac{2+1}{2+1+2+1}\right)\left(\frac{3+1}{3+1+1+1}\right)$$

not smoothed

$$= \frac{4}{10} \cdot \frac{3}{6} \cdot \frac{4}{6} = \frac{4}{30}$$

$$P(\text{neutral} \mid f, M) \propto P(\text{neutral}) \cdot P(f \mid \text{neutral}) \cdot P(M \mid \text{neutral})$$

$$= \left(\frac{1}{10}\right)\left(\frac{0+1}{0+1+1+1}\right)\left(\frac{1+1}{1+1+0+1}\right)$$

not smoothed!

$$= \left(\frac{1}{10}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{2}{90} = \frac{4}{180}$$

$P(\text{good} \mid f, M)$ has the largest numerator,

$\therefore$ we predict <u>good</u>.

Follow along with the demo by clicking the **code** link on the course website next to Lecture 16.

Summary

Summary

- In classification, our goal is to predict a discrete category, called a **class**, given some features.
- The Naive Bayes classifier works by estimating the numerator of $P(\text{class}|\text{features})$ for all possible classes.
- It uses Bayes' theorem:

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

- It also uses a simplifying assumption, that features are conditionally independent given a class:

$$P(\text{features}|\text{class}) = P(\text{feature}_1|\text{class}) \cdot P(\text{feature}_2|\text{class}) \cdot \ldots$$

Next time

- Next time, we'll look at another practical use case for Naive Bayes — text classification.
- Time permitting, we'll briefly cover a different classification technique — logistic regression.
  - Will not be on the exam.
- Lecture 18 (the final lecture) and the final groupwork will consist solely of review.