

Module 23 – Naive Bayes



DSC 40A, Summer 2023

Agenda

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

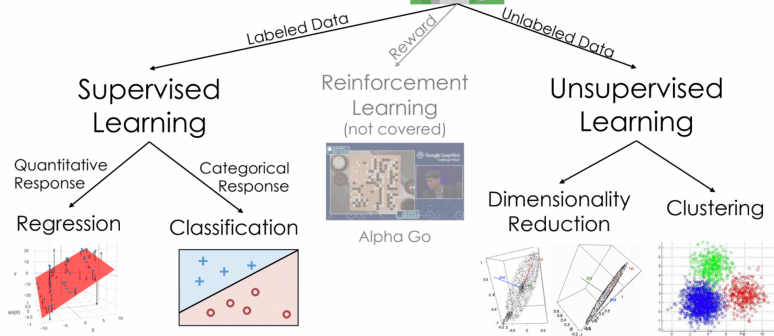
Recap: Bayes' theorem, independence, and conditional independence

- ▶ 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.

Classification

Taxonomy of machine learning

Taxonomy of Machine Learning



Classification problems

- ▶ Like with regression, we're interested in making predictions based on data (called **training data**) for which we know the value of the response variable.
- ▶ The difference is that the response variable is now **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?

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.

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

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

Example: avocados

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

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

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

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:

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

- ▶ 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

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

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

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

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

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.

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

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

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?

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.

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

- ▶ 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

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

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

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

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?

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

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?

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

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!

Dictionary

Definitions from [Oxford Languages](#) · [Learn more](#)



na·ive

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: avocados, again

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

Uh oh...

- ▶ There are no soft unripe avocados in the data set.
- ▶ The estimate $P(\text{soft}|\text{unripe}) \approx \frac{\# \text{ soft unripe avocados}}{\# \text{ unripe avocados}}$ is 0.
- ▶ The estimated numerator,
 $P(\text{unripe}) \cdot P(\text{soft, green-black, Hass}|\text{unripe}) = P(\text{unripe}) \cdot P(\text{soft}|\text{unripe}) \cdot P(\text{green-black}|\text{unripe}) \cdot P(\text{Hass}|\text{unripe})$,
is also 0.
- ▶ But just because there isn't a soft unripe avocado in the data set, doesn't mean that it's impossible for one to exist!
- ▶ **Idea:** Adjust the numerators and denominators of our estimate so that they're never 0.

Smoothing

- ▶ **Without** smoothing:

$$P(\text{soft}|\text{unripe}) \approx \frac{\# \text{ soft unripe}}{\# \text{ soft unripe} + \# \text{ medium unripe} + \# \text{ firm unripe}}$$

$$P(\text{medium}|\text{unripe}) \approx \frac{\# \text{ medium unripe}}{\# \text{ soft unripe} + \# \text{ medium unripe} + \# \text{ firm unripe}}$$

$$P(\text{firm}|\text{unripe}) \approx \frac{\# \text{ firm unripe}}{\# \text{ soft unripe} + \# \text{ medium unripe} + \# \text{ firm unripe}}$$

- ▶ **With** smoothing:

$$P(\text{soft}|\text{unripe}) \approx \frac{\# \text{ soft unripe} + 1}{\# \text{ soft unripe} + 1 + \# \text{ medium unripe} + 1 + \# \text{ firm unripe} + 1}$$

$$P(\text{medium}|\text{unripe}) \approx \frac{\# \text{ medium unripe} + 1}{\# \text{ soft unripe} + 1 + \# \text{ medium unripe} + 1 + \# \text{ firm unripe} + 1}$$

$$P(\text{firm}|\text{unripe}) \approx \frac{\# \text{ firm unripe} + 1}{\# \text{ soft unripe} + 1 + \# \text{ medium unripe} + 1 + \# \text{ firm unripe} + 1}$$

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

Example: avocados, with smoothing

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 soft green-black Hass avocado. Using Naive Bayes, **with smoothing**, would you predict that your avocado is ripe or unripe?

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:

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