Lecture 17

# Naïve Bayes

DSC 40A, Spring 2024

# Announcements

- Homework 7 is due **tonight**. **New: You can use two slip days on it.**
- Homework 8, the final homework, will be released tomorrow and will be due on Thursday, June 6th. **New: You cannot use slip days on it, but it'll be max 3 questions.**
- Make sure you've watched the recorded lecture from Tuesday and read the accompanying lecture note.
- Look at the solutions to last Monday's groupwork worksheet posted on Ed!
- Read the new Advice page written by the tutors.

# The Final Exam is on Saturday, June 8th!

- The Midterm Exam is on **Saturday, June 8th from 8-11AM**.
  - You will receive a randomized seat assignment early next week.
- 180 minutes, on paper, no calculators or electronics.
  - **You are allowed to bring two double-sided index cards (4 inches by 6 inches) of notes that you write by hand (no iPad).**
- Content: All lectures (including next week), homeworks, and groupworks.
- We will have two review sessions. In each of them, the first hour will be a mock exam **which you will take silently on paper**; we will take up the problems in the second half.
  - Tuesday, June 4th, 5-7PM (empirical risk minimization and linear algebra).
  - Thursday, June 6th, 5-7PM (gradient descent and probability).
- Friday, June 7th, 4-9PM: office hours in HDSI 123.
- Prepare by practicing with old exam problems at practice.dsc40a.com.

# Agenda

- Classification.
- Classification and conditional independence.
- Naïve Bayes.

# Recap: Bayes Theorem', independence, and conditional independence

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

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

- $A$ and $B$ are **independent** if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

- $A$ and $B$ are **conditionally independent** given $C$ if:

$$\mathbb{P}((A \cap B)|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$$

    - In general, there is no relationship between independence and conditional independence.

# Question 🤔

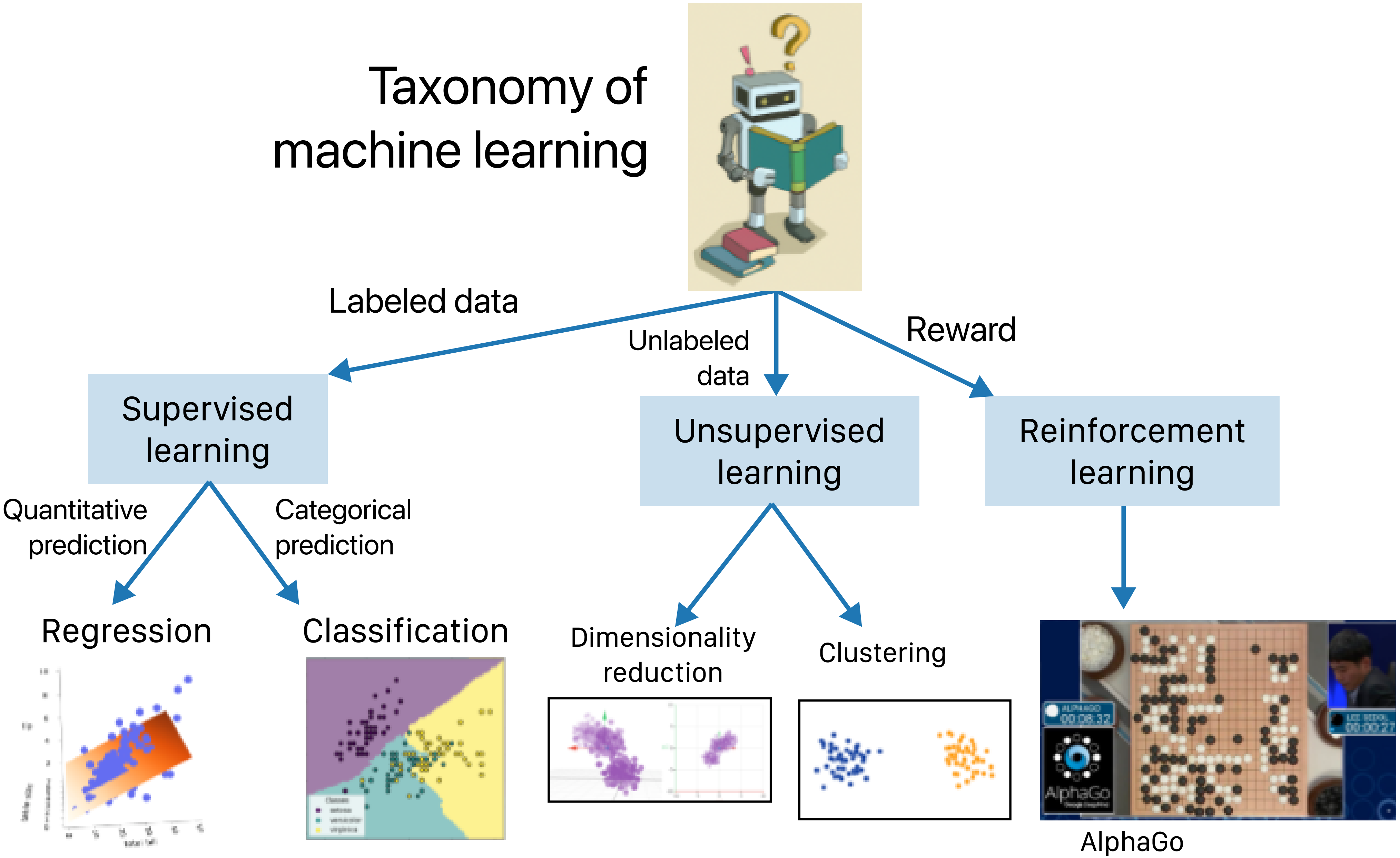
**Answer at q.dsc40a.com**

**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.

# Classification

# 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.
  - Predicting whether you'll be late to school or not.

# Example: Avocados

| 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, and want to know if it is ripe.

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

# Example: Avocados

| 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, and want to know if it is ripe. Based on this data, would you predict your avocado is ripe or unripe?

**Strategy**: Calculate two probabilities:

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

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

## Estimating probabilities

- We would like to determine $\mathbb{P}(\text{ripe}|\text{green-black})$ and $\mathbb{P}(\text{unripe}|\text{green-black})$ 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.

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

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

# Example: Avocados

| 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, and want to know if it is ripe. Based on this data, would you predict your avocado is ripe or unripe?

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

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

## 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:

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

- More generally:

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

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

# Example: Avocados

| 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, and want to know if it is ripe. Based on this data, would you predict your avocado is ripe or unripe?

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

# Example: Avocados

| 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, and want to know if it is ripe. Based on this data, would you predict your avocado is ripe or unripe?

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

# Example: Avocados

| 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, and want to know if it is ripe. Based on this data, would you predict your avocado is ripe or unripe?

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

**Shortcut**: Both probabilities have the same denominator, so the larger probability is the one with the **larger numerator**.

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

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

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

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

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

# 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?

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

**Issue**: We have not seem a firm green-black Zutano avocado before, which means that the following probabilities are undefined:

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

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

# A simplifying assumption

- We want to find $\mathbb{P}(\text{ripe} | \text{firm, green-black, Zutano})$, but there are no firm green-black Zutano avocados in our dataset.
- Bayes' Theorem tells us this probability is equal to:

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

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

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

# 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?

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

# 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?

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

## Conclusion

- The numerator of $\mathbb{P}(\text{ripe}|\text{firm, green-black, Zutano})$ is $\frac{6}{539}$.
- The numerator of $\mathbb{P}(\text{unripe}|\text{firm, green-black, Zutano})$ is $\frac{6}{88}$.
- Both probabilities have the same denominator, $\mathbb{P}(\text{firm, green-black, Zutano})$.
- 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** ❌.

# Naïve Bayes

# The Naïve Bayes classifier

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

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

- For each class, we compute the numerator using the **naïve 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

/näˈēv/

*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 $\mathbb{P}(\text{soft}|\text{unripe}) \approx \frac{\#\text{ soft unripe avocados}}{\#\text{ unripe avocados}}$ is 0.
- The estimated numerator:

  $$\mathbb{P}(\text{unripe}) \cdot \mathbb{P}(\text{soft, green-black, Hass}|\text{unripe}) = \mathbb{P}(\text{unripe}) \cdot \mathbb{P}(\text{soft}|\text{unripe}) \cdot \mathbb{P}(\text{green-black}|\text{unripe}) \cdot \mathbb{P}(\text{Hass}|\text{unripe})$$

  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:

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

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

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

- **With** smoothing:

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

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

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

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

# 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. Based on this data, 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 Naïve Bayes classifier uses Bayes' Theorem:

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

- And works by estimating the numerator of $\mathbb{P}(\text{class}|\text{features})$ for all possible classes.
- It also uses a simplifying assumption, that features are conditionally independent given a class:

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