Lecture 17

# Naïve Bayes

DSC 40A

## 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.

new $\mathbb{P}$ old $\mathbb{P}$

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

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

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

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

Equivalent

- $A$ 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 🤔

Take a moment to pause and reflect...

If you have any questions please post online to our forms/Q&A site.

Course staff will answer them ASAP!

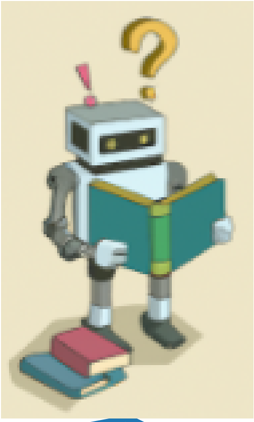
# Classification

we Know what they ARE!

data labels

X y

# Taxonomy of machine learning



data ~~labels~~

X ~~y~~

Labeled data

Unlabeled data

Reward

Supervised learning

Unsupervised learning

Reinforcement learning

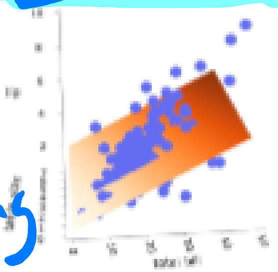
Quantitative prediction

Categorical prediction

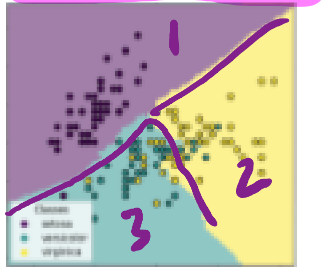
classes e.g. "good" / "bad"
1 / 0

Regression

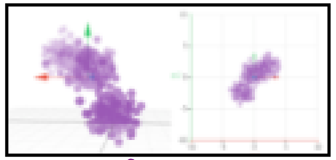
Continous values e.g. Min Driving



Classification



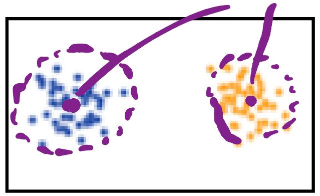

Dimensionality reduction



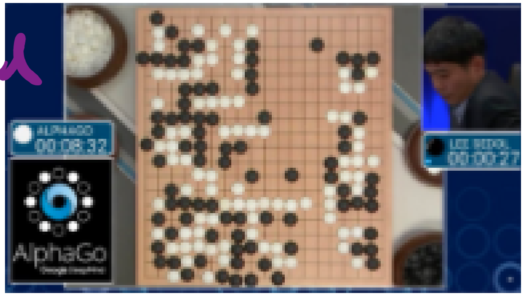
PCA

Clustering

centroid



K means



AlphaGo

# 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. ⟹ $y_i$

labels

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- The difference is that the response variable is now **categorical**.
- Categories are called **classes**.
- Example classification problems:
  - yes or NO — Deciding whether a patient has kidney disease.
  - Identifying handwritten digits.

    y: two two two . three

    ~~X~~: 2 2 2 2 3

    0 → 9

    10 classes
  - Determining whether an avocado is ripe.
  - Predicting whether credit card activity is fraudulent.
  - Predicting whether you'll be late to school or not.

Regression $\vec{y}$ $\begin{bmatrix} 22 \\ 31.5 \\ 20 \\ \vdots \end{bmatrix}$

Classification $\vec{y}$ $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$

0

1

82% yes

18% no

⟹ yes

yes or no

~~maybe so?~~

Training Data

# Example: Avocados

$x_1$ $y$

| 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 ✅ |

R
NR
R
R
NR

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

11 examples "observations"
5 G.B. Avo.

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

3 R > 2 NR

Our G.B.A. = Ripe

# Example: Avocados

$H(G.B.A.) = $ "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 ✅ |

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:

$$\mathbb{P}(\text{ripe}|\text{green-black}) = \frac{3}{5}$$

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

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

Population Parameter !!

## 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. Sample statistic

Guess or Infer Population Param

$$\mathbb{P}(\text{ripe}|\text{green-black}) \approx \frac{\#\text{ ripe green-black avocados in sample}}{\#\text{ green-black avocados in sample}} = \frac{3}{5}$$

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

# Example: Avocados

$x_n$ $x_1$ $y$

| 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?

in ML not typically 1 feature, but many!

$$\mathbb{P}(\text{ripe}|\text{green-black}) = \frac{3}{5}$$

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

$\mathbb{P}(B|A) = \mathbb{P}(\text{ripe} \mid \text{G.B.A.})$

# 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:

CLASS = {R, NR}

CLASS

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

feature

$\mathbb{P}(B) = \mathbb{P}(\text{ripe})$

$\mathbb{P}(A) = \mathbb{P}(\text{G.B.A.})$

$\mathbb{P}(A|B) = \mathbb{P}(\text{GBA} \mid \text{ripe})$

- 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?

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

$$\mathbb{P}(\text{ripe}|\text{GBA}) = \frac{\mathbb{P}(\text{ripe}) \cdot \mathbb{P}(\text{G-BA}|\text{Ripe})}{\mathbb{P}(\text{GBA})}$$

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

# Example: Avocados

| color | ripeness |
|---|---|
| bright green | unripe ❌ |
| green-black | ripe ✅ |
| purple-black | ripe ✅ |
| ① green-black | unripe ❌ |
| purple-black | ripe ✅ |
| bright green | unripe ❌ |
| green-black | ripe ✅ |
| purple-black | ripe ✅ |
| green-black | ripe ✅ |
| ② 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?

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

$$\mathbb{P}(\text{unripe}|\text{GBA}) = \frac{\mathbb{P}(\text{unripe}) \cdot \mathbb{P}(\text{GBA}|\text{unripe})}{\mathbb{P}(\text{GBA})}$$

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

# 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**.

"Proportional to"

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

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

larger still

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

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

# 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 |

$X$ $y$

You have a $\underbrace{\text{firm}}_{x_1}$ $\underbrace{\text{green-black}}_{x_2}$ $\underbrace{\text{Zutano}}_{x_3}$ avocado. Based on this data, would you predict that your avocado is ripe or unripe?

GB, F, Z

# 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.

$$\mathbb{P}(\text{ripe} | \text{firm, green-black, Zutano}) = 0$$

$$\mathbb{P}(\text{unripe} | \text{firm, green-black, Zutano}) = 0$$

$\frac{0}{0}$ ? WTH!

# 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:

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

?

ML is all about Assumptions ½ Joke!

## A simplifying assumption

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

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

P(class | features)

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

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

# 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?

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

$$\propto P(\text{ripe}) \cdot P(\text{Firm}, \text{GBA}, Z \mid \text{ripe})$$

$$= P(\text{ripe}) \cdot P(\text{Firm} \mid \text{ripe}) \cdot P(\text{GBA} \mid \text{ripe}) \cdot P(Z \mid \text{ripe})$$

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

Assume the Features are Cond. Ind. given the class ripe || 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?

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

$$\propto \mathbb{P}(UN) \cdot \mathbb{P}(F|UN) \cdot \mathbb{P}(G\text{-}BA|UN) \cdot \mathbb{P}(Z|UN)$$

$$= \frac{\cancel{4}}{11} \cdot \frac{3}{\cancel{4}} \cdot \frac{2}{4}\,\frac{1}{2} \cdot \frac{2}{4}\,\frac{1}{2} = \frac{3}{44}$$

$$\frac{6}{539} \text{ vs. } \frac{3}{44} \rightarrow \frac{6}{88}$$

# 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** ❌.

Larger

$H(\text{firm}, \text{GBA}, \text{Zutano}) = \text{unripe}$

# Naïve Bayes

# The Naïve Bayes classifier

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

Diff

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

- 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.

4 7

UN = unripe

ripe vs. unripe

# Example: Avocados, again

S G.B. H

| 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?

$P(\text{ripe} \mid \text{soft}, GB, \text{Hass}) \rightarrow$ expand out via conditional assumption

$\propto P(\text{ripe}) \cdot P(s \mid \text{ripe}) \cdot P(GB \mid \text{ripe}) \cdot P(h \mid \text{ripe})$

$\frac{7}{11} \cdot \frac{4}{7} \cdot \frac{3}{7} \cdot \frac{5}{7} =$ Something

$\propto P(\text{unripe}) \cdot P(s \mid UN) \cdot P(GB \mid UN) \cdot P(h \mid UN)$

$\frac{\cancel{4}}{11} \cdot \frac{0}{\cancel{4}} \cdot \frac{2}{4} \frac{1}{2} \cdot \frac{2}{4} \frac{1}{2} \boxed{= 0}$

# 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:

  no smoothing as it IS NOT Cond.

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

  +1 yes

  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.

Just ADD 1

## Smoothing

- **Without** smoothing:

No occurances in Data

$$\mathbb{P}(\text{soft}|\text{unripe}) \approx \frac{\#\text{ soft unripe}}{\#\text{ soft unripe} + \#\text{ medium unripe} + \#\text{ firm unripe}} \Rightarrow 0$$

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

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

- **With** smoothing:

+3 because we have 3 categories of "softness"

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

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

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

- Smoothing: +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?

$$\mathbb{P}(\text{ripe} \mid S, GB, H)$$

$$\propto \mathbb{P}(\text{ripe}) \cdot \mathbb{P}(S \mid \text{ripe}) \cdot \mathbb{P}(GB \mid \text{ripe}) \cdot \mathbb{P}(H \mid \text{ripe})$$

$$\frac{7}{11} \cdot \frac{4+1}{7+3} \cdot \frac{3+1}{7+3} \cdot \frac{5+1}{7+2}$$

{soft, med. firm} {GB, PB, BG} {HASS, Zutano}

Do not Smooth!

$$= \frac{7}{11} \cdot \frac{\cancel{5}}{\cancel{10}_{5}} \cdot \frac{\cancel{4}^{2}}{\cancel{10}_{5}} \cdot \frac{\cancel{6}^{2}}{\cancel{9}_{3}} = \frac{14}{165}$$

$11 \cdot 5 \cdot 3$

# 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?

$$\mathbb{P}(\text{unripe} \mid S, GB, H)$$

$$\propto \mathbb{P}(UN) \cdot \mathbb{P}(S \mid UN) \cdot \mathbb{P}(GB \mid UN) \cdot P(H \mid UN)$$

$$= \frac{4}{11} \cdot \frac{0+1}{4+3} \cdot \frac{2+1}{4+3} \cdot \frac{2+1}{4+2}$$

$$= \frac{\cancel{4}\,2}{11} \cdot \frac{1}{7} \cdot \frac{\cancel{3}}{7} \cdot \frac{3}{\cancel{6}\,\cancel{2}} = \frac{6}{539}$$

ripe numerator $= \frac{14}{165}$ = LARGER

unripe numerator $= \frac{6}{539}$

given (Soft, G.B, HASS) Avocado

Predict $\boxed{\text{ripe}}$

# 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:

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