Lecture 23 – Naive Bayes

DSC 40A, Spring 2023

Announcements

- Homework 7 is released, due **Thursday 6/1 at 11:59pm**.
- Monday is a holiday. No lecture, and no office hours.
- **Midterm 2 is Monday 6/5** during lecture.
  - You'll be allowed an unlimited number of handwritten note sheets for Midterm 2 (and Final Part 2). Start studying and preparing your notes now!
  - Midterm 2 covers Homeworks 5 through 7. Clustering is included, but the vast majority will be probability and combinatorics.

Agenda

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

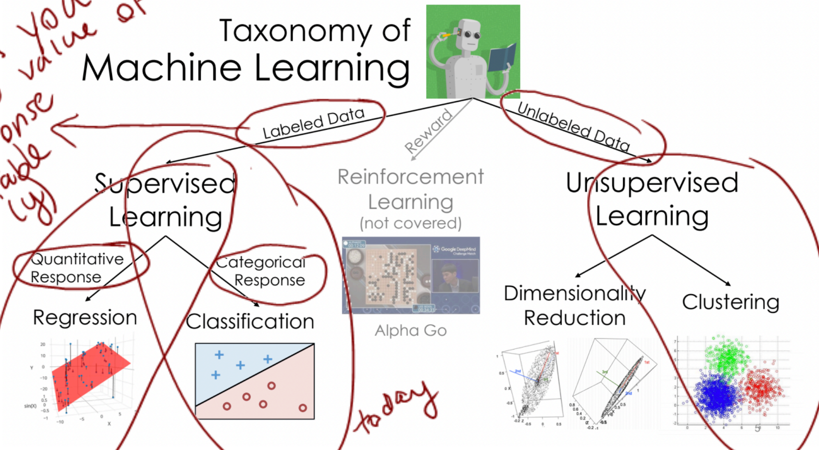
Recap: Bayes' theorem, independence, and conditional independence

- Bayes' theorem: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
- A and B are **independent** if $P(A \cap B) = P(A) \cdot P(B)$. (one of 3 equivalent definitions)
- 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

means you know value of response variable (y)



today

y = Salary

y = ripe/unripe

no labels

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

new

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

color	ripeness
bright green	unripe
green-black	ripe ①
purple-black	ripe
green-black	unripe ①
purple-black	ripe
bright green	unripe
green-black	ripe ②
purple-black	ripe
green-black	ripe ③
green-black	unripe ②
purple-black	ripe

training data

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

ripe – 3 out of 5 green-black avocados are ripe

unripe – 2 out of 5 green-black avocados are 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:

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

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

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

likely ripe because $\frac{3}{5} > \frac{2}{5}$

Estimating probabilities

population parameter

- 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.

$$\underbrace{P(\text{ripe}|\text{green-black})}_{\text{parameter}} \approx \underbrace{\frac{\#\text{ ripe green-black avocados in sample} \;{}^{=3}}{\#\text{ green-black avocados in sample} \;{}_{=5}}}_{\text{estimate based on sample}}$$

- 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

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

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

direct interpretation

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:

$$P(\overset{\text{ripe}}{\text{class}}|\overset{\text{green-black}}{\text{features}}) = \frac{P(\overset{\text{ripe}}{\text{class}}) \cdot P(\overset{\text{green-black}}{\text{features}}|\overset{\text{ripe}}{\text{class}})}{P(\underset{\text{green-black}}{\text{features}})}$$

- 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

Using Bayes Thm

$$P(\underbrace{\text{class}}_{\text{ripe}}|\underbrace{\text{features}}_{\text{green-black}}) = \frac{P(\text{class})\cdot P(\text{features}|\text{class})}{P(\text{features})}$$

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

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

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

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

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

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

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?

means "proportional to"

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.

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

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

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

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

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

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

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

new

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

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

shortcut to avoid having to calculate denominator

- **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

proportional to: numerator only

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

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

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

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

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?

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

$$\propto \underbrace{P(\text{unripe})}_{\frac{\cancel{4}}{11}} \cdot \underbrace{P(\text{firm}|\text{unripe})}_{\frac{3}{\cancel{4}}} \cdot \underbrace{P(\text{green-black}|\text{unripe})}_{\frac{\cancel{2}}{\cancel{4}}} \cdot \underbrace{P(\text{zutano}|\text{unripe})}_{\frac{\cancel{2}}{4}} = \frac{3}{44} = \boxed{\frac{6}{88}}$$

Conclusion

tiny

- The numerator of $P(\text{ripe}|\text{firm, green-black, Zutano})$ is $\boxed{\frac{6}{539}}$.
- The numerator of $P(\text{unripe}|\text{firm, green-black, Zutano})$ is $\boxed{\frac{6}{88}}$.

bigger

  - 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

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

ripe/unripe

- 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

new

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

$$P(\text{ripe} \mid \text{soft}, \text{green-black}, \text{Hass}) \propto P(\text{ripe}) \cdot P(\text{soft}, \text{green-black}, \text{Hass} \mid \text{ripe})$$

naive by assumption

$$= P(\text{ripe}) \cdot P(\text{soft} \mid \text{ripe})\, P(\text{green-black} \mid \text{ripe}) \cdot P(\text{Hass} \mid \text{ripe})$$

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

$$P(\text{unripe} \mid \text{soft}, \text{green-black}, \text{Hass}) \propto P(\text{unripe}) \cdot P(\text{soft}, \text{green-black}, \text{Hass} \mid \text{unripe})$$

naive by assumption

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

$$= \frac{4}{11} \cdot \frac{0}{4} \cdots = 0$$

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:

add to one

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

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

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

- **With** smoothing:

add to one

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

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

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

- 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?

Smoothing

	ripe	unripes
soft		
med		
firm		

- **Without** smoothing:

add to 1

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

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

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

- **With** smoothing:

still add to 1

add 1 to top, 3 to bottom

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

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

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

because 3 levels of firmness

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

Example: avocados, with smoothing

without smoothing : $\frac{4}{7}$

with smoothing : $\frac{5}{10}$

P(Hass|ripe) without smoothing : $\frac{\#\text{ripe Hass}}{\#\text{ripe Hass} + \#\text{ripe Zutano}}$

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?

$$P(\text{ripe}|\text{soft},\text{gb},\text{Hass}) \propto P(\text{ripe}) \cdot P(\text{soft},\text{gb},\text{Hass}|\text{ripe})$$
$$= P(\text{ripe}) \cdot \boxed{P(\text{soft}|\text{ripe})} \cdot P(\text{gb}|\text{ripe}) \cdot \boxed{P(\text{Hass}|\text{ripe})}$$
$$= 7/11 \cdot 5/10 \cdot 4/10 \cdot 6/9$$

$$P(\text{unripe}|\text{soft},\text{gb},\text{Hass}) \propto P(\text{unripe}) \cdot P(\text{soft},\text{gb},\text{Hass}|\text{unripe})$$
$$= P(\text{unripe}) \cdot P(\text{soft}|\text{unripe}) \cdot P(\text{gb}|\text{unripe}) \cdot P(\text{Hass}|\text{unripe})$$
$$= 4/11 \cdot 1/7 \cdot 3/7 \cdot 3/6$$

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:

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