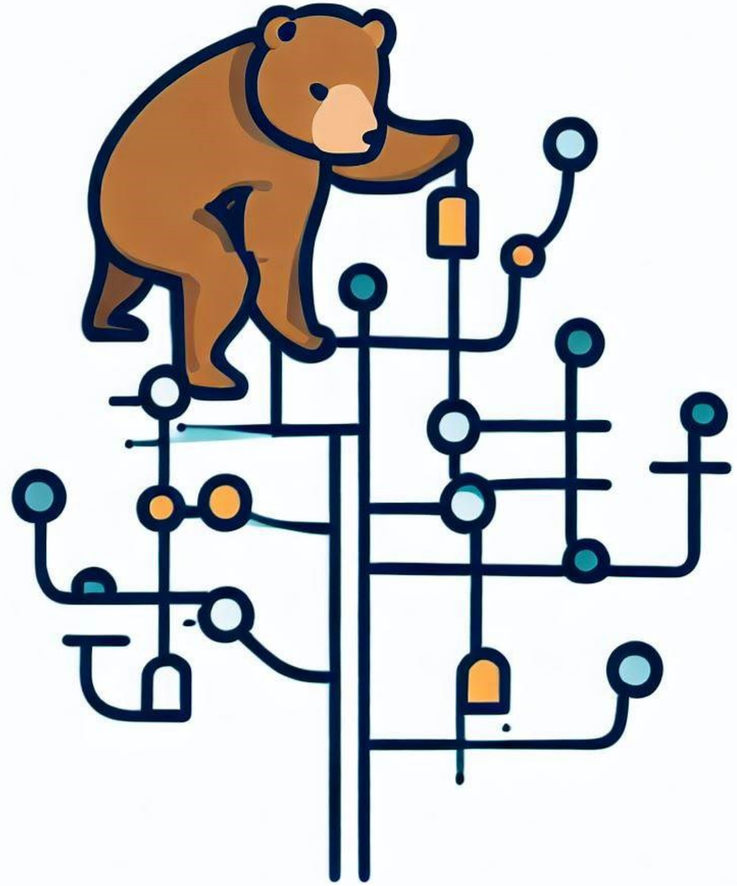
Advanced Techniques

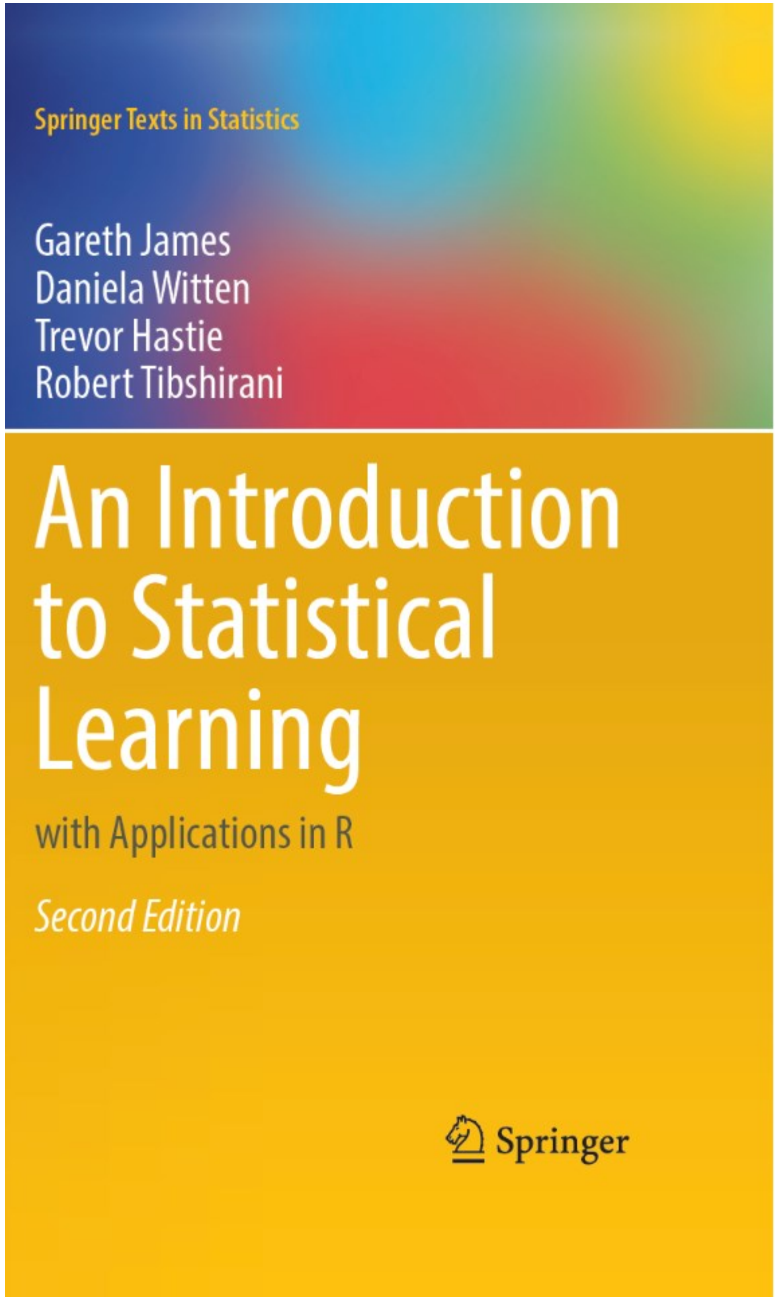
UCSB SUMMER STATS WORKSHOP 2023

WEEK 6



Plan for Today

1. Dan: decision trees
   - Decision tree exercise
2. Ashley: cluster analysis
   - Cluster analysis exercise
3. Final project help



https://www.statlearning.com/

Decision trees

Supervised machine learning technique

Attempts to predict an outcome variable using some set of predictor variables

```
outcome ~ predictor1 + predictor2
```

Normally do this with some kind of linear model:

- outcome = B_1*predictor1 + B_2*predictor2 + B_3*(predictor1*predictor2) + B_0

Main effects **Interaction** **Intercept**

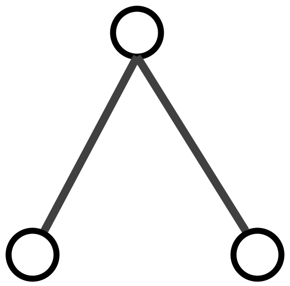
Decision trees

Recursive binary partitioning

Decision trees

Recursive **binary partitioning**

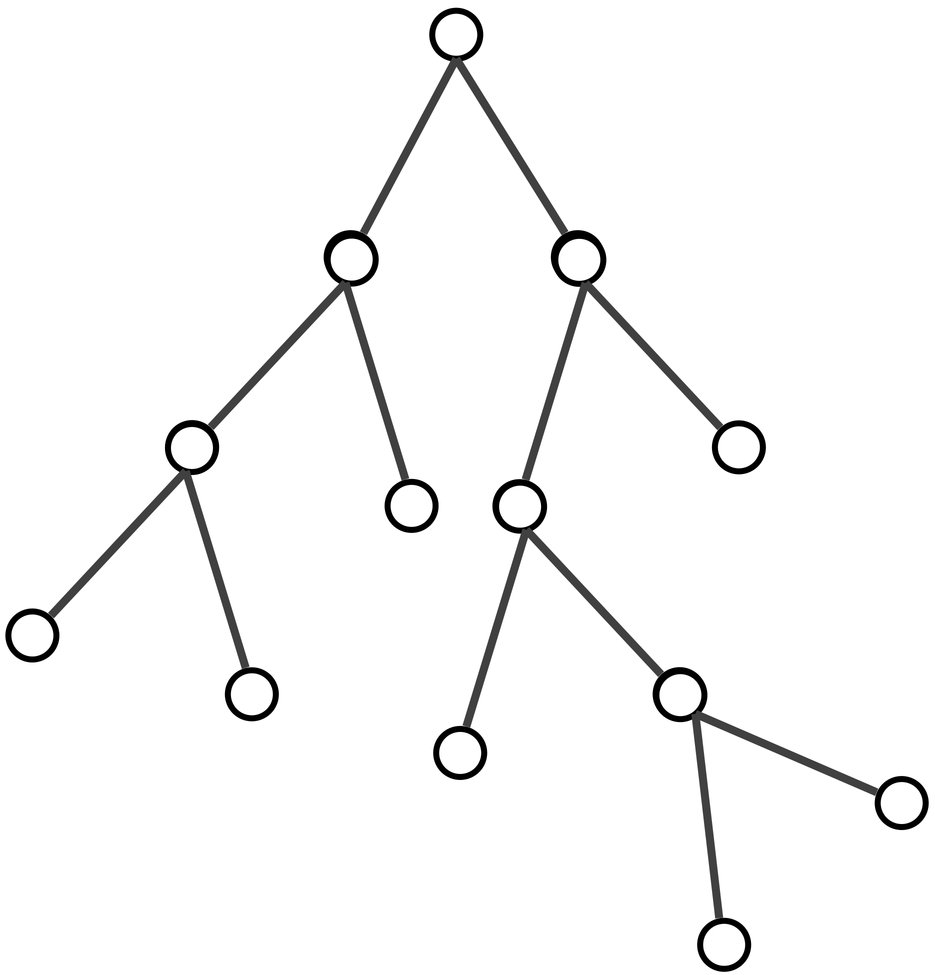
- Split the data into two parts…
- …according to one of the predictor variables…
- …usually so as to:
  - Maximize differences *between* partitions and/or
  - Minimize differences *within* partitions



Decision trees

Recursive binary partitioning

- Split the splits



Suppose you're going on vacation…

Need to decide what clothes to pack

Two possible cities:

- New York City
- Santa Barbara

Two times of year:

- Winter
- Summer

Collect weather data from Wikipedia

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

Suppose you're going on vacation…

ANOVA approach:

vacayAnova <- lm(temp ~ season * city)

Probably a significant interaction between city and season in predicting temperature

But how does that help you pack?

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1

M = 61.78, SD = 5.84

City	Season	Av. Temp (*F*)
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

M = 54.78, SD = 23.19

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
New York City	Summer (June)	72
New York City	Summer (July)	77.5

M = 70.73, SD = 5.32

City	Season	Av. Temp (*F*)
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

M = 45.83, SD = 12.76

City partition: ✗

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1

M = 61.78, SD = 5.84

City	Season	Av. Temp (*F*)
New York City	Summer (June)	72
New York City	Summer (July)	77.5
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

M = 54.78, SD = 23.19

Season partition: ✓

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
New York City	Summer (June)	72
New York City	Summer (July)	77.5

M = 70.73, SD = 5.32

City	Season	Av. Temp (*F*)
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

M = 45.83, SD = 12.76

City	Season	Av. Temp (*F*)
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

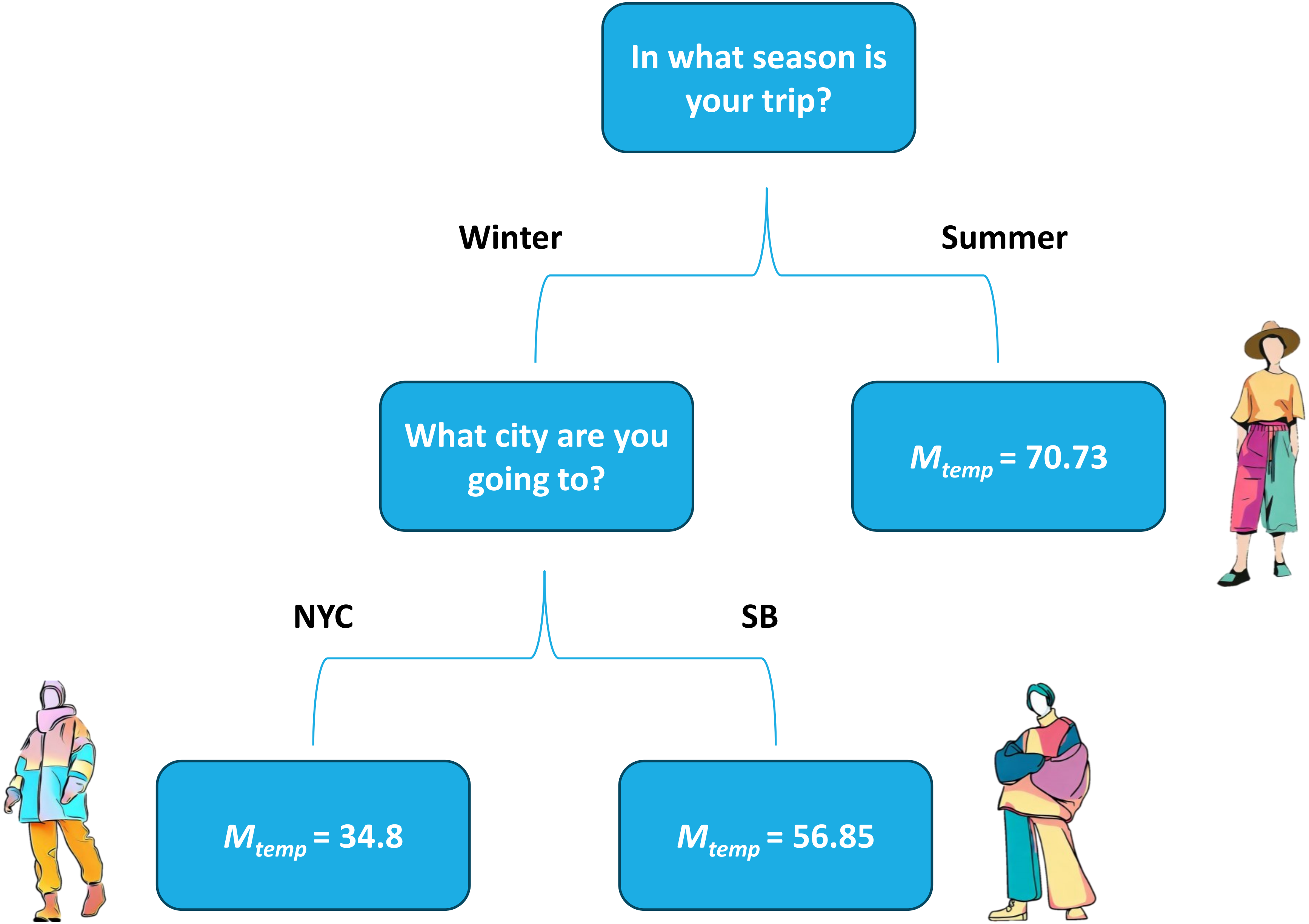
City	Season	Av. Temp (*F*)
Santa Barbara	Winter (Jan.)	56.6
Santa Barbara	Winter (Feb.)	57.1

City	Season	Av. Temp (*F*)
New York City	Winter (Jan.)	33.7
New York City	Winter (Feb.)	35.9

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3
New York City	Summer (June)	72
New York City	Summer (July)	77.5

City	Season	Av. Temp (*F*)
Santa Barbara	Summer (June)	65.1
Santa Barbara	Summer (July)	68.3

City	Season	Av. Temp (*F*)
New York City	Summer (June)	72
New York City	Summer (July)	77.5


In what season is your trip?
Winter
Summer
What city are you going to?
M_{temp} = 70.73
NYC
SB
M_{temp} = 34.8
M_{temp} = 56.85

Decision trees

Straightforward to interpret

Immediately lend themselves to decision making

Especially useful in industry contexts

- But also good for science

Zoe Liberman

Social cognitive development, food, language

https://liberman.psych.ucsb.edu/research

Example Data

n = 400 children

- Ages 3-12

Children watched a video of two dolls interacting

One doll (**the sharer**) shared food with a second doll (**the recipient**)

Sharer either shared:

- A liked food (cookies)
- A disliked food (broccoli)

Also shared either:

- 1 piece of food
- 5 pieces of food

Afterwards, children were given the choice to play with either doll

- Recorded which doll the children chose

Additionally, parents rated children using a 7-point Likert scale on:

- Pickiness about vegetables (1 = not picky at all, 7 = extremely picky)
- Pickiness about fruits (1 = not picky at all, 7 = extremely picky)

Example Data: Decision Trees

RQ1: Do children prefer a doll who shares a liked food over a disliked food?

RQ2: Do children prefer a doll who shares more over a doll who shares less?

RQ3: Do attitudes about sharing differ with age?