

# CS260: Machine Learning Theory

## Lecture 1: Course Introduction

Jenn Wortman Vaughan

September 26, 2011

What is machine learning?

# What is machine learning?

Machine learning is the study of how to use past observations or experience to **automatically and efficiently learn** to make better predictions or choose better actions in the future

# Movie Recommendations

The screenshot shows a user interface for a movie recommendation system. On the left, there is a card for '30 Rock: Season 2' featuring a cast photo and a 5-star rating. A red tooltip box is overlaid on the right side of the card, providing detailed information about the season. The tooltip includes the title, year, rating, and number of discs/episodes. It also contains a synopsis, a list of starring actors and the director, the genre, and the format. At the bottom of the tooltip, there is a 4.7 star rating and a recommendation based on the user's interests in other shows.

**Play** **In Q**

★★★★★

**30 Rock: Season 2**

2007 **NR** 2 discs / 15 episodes

The second season of this Emmy-winning NBC sitcom -- written by funnywoman Tina Fey of "Saturday Night Live" fame, who also stars -- picks up where the show's tumultuous first season of laughs left off.

**Starring:** Tina Fey, Alec Baldwin

**Director:** Don Scardino

**Genre:** TV Sitcoms

**Format:** DVD and streaming (HD available)

★★★★★ **4.7** Our best guess for Jenn

♥ Recommended based on your interest in *Office Space*, *Weeds: Season 3* and *The Office: Season 5*

**30 Rock: Season 4**

# Click Prediction

furniture stores in la - Google Search

http://www.google.com/

Apple Yahoo! Google Maps YouTube Wikipedia News (240) Popular

Web Images Videos Maps News Shopping Gmail more

Web History | Search settings | Sign in

## Google

furniture stores in LA

About 707,000 results (0.40 seconds)

Instant is on

Advanced search

**Everything**

- Images
- Videos
- Maps
- More

Show search tools

**Los Angeles Furniture**  
LivingSpaces.com/LosAngeles The styles you love at prices you can afford. Same day delivery.

**Find Discount Furniture**  
www.EasyLifeFurniture.com Discount Furniture Stores in Four Orange County Locations

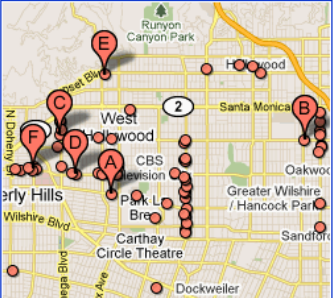
**Modern Furniture Store LA**  
www.modani.com Modern furniture at wholesale price Grand Opening Get 50% off!

**LA Furniture Store**  
We ship nationwide. Huge variety in modern furniture, contemporary & Italian furniture like platform bed, leather sofa, sectional sofas & bedroom furniture ...  
Modern Sofas - Bedroom - Living Room - Dining Room  
www.lafurniturestore.com/ - Cached - Similar

**Los Angeles Furniture Stores in Los Angeles CA Yellow Pages by ...**  
Directory of Los Angeles Furniture Stores in CA yellow pages. Find Furniture Stores in Los Angeles maps with reviews, websites, phone numbers, addresses, ...  
www.superpages.com/.../C-Furniture+Stores/.../T-Los+Angeles/ - Cached - Similar

**Los Angeles Furniture Store | Furniture Los Angeles Discount**  
Furniture Discount Store Los Angeles. Visit our Showroom Furniture in Los Angeles and see Sofas, Beds, Dining, Mattresses, Kids, Home Office, ...  
www.furniturevision.com/ - Cached - Similar

**Local business results for furniture stores in la near West Hollywood, CA** - Change location



- A Mitchell Gold LA**  
www.mitchellgoldla.com - (323) 651-0200 - More
- B Edmon's Unique Furniture and Stone Gallery**  
www.edmonsgallery.com - (323) 462-5787 - More
- C Docantic inc**  
www.morateur.com - (310) 360-0588 - More
- D Blueprint Furniture**  
www.blueprintfurniture.com - (323) 653-2439 - 51 reviews
- E CB2**  
www.cb2.com - (323) 848-7111 - 3 reviews
- F Diva Furniture**  
www.divafurniture.com - (310) 278-3191 - 7 reviews

**Sponsored links**

**Buy Furniture Direct**  
Don't pay retail! Buy furniture direct from top manufacturers.  
www.DirectHomeDiscount.com

**Furniture in Los Angeles**  
4 Huge Los Angeles Furniture Stores Top Rated on Yelp and Citysearch  
www.thesofaco.com  
Thesofaco.com is rated ★★★★★

**40-60% Off Furniture**  
Enjoy Huge Savings On Our Quality Furniture. For A Limited Time Only.  
JCPenney.com/FurnitureSale  
JCPenney.com is rated ★★★★★

**Furniture in Los Angeles**  
Quality furniture in LA. Free delivery! Unbeatable prices.  
www.Furniture2Go.com

**Pottery Barn® Furniture**  
Decorate Your Home with Quality & Stylish Furniture from Pottery Barn  
www.PotteryBarn.com

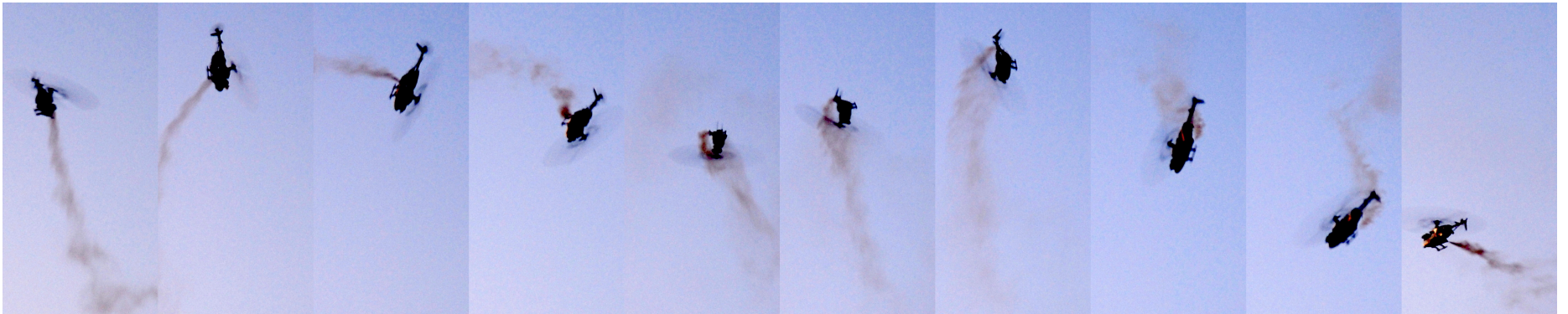
**Quality Furniture 4 Less**  
We guarantee the lowest price Come by one of our showrooms  
www.astarfurn.com  
Los Angeles, CA

**Discount Furniture**  
Free Delivery In So. California Save 50-70% off Retail Prices  
www.ebudgetfurniture.com

**La-Z-Boy® - Official Site**  
Visit La-Z-Boy for Comfort & Style Recliners, Sofas, Sleepers & More.  
www.La-Z-Boy.com

# Autonomous Flight

Helicopter rolls:



Helicopter flips:



# Other Examples

- Medical diagnosis
- Handwritten character recognition
- Customer segmentation (marketing)
- Document segmentation (classifying news)
- Spam filtering
- Weather prediction and climate tracking
- Gene prediction
- Face recognition

# Spam Prediction

We are given a set of labeled email messages

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight ✓

To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting ✓

To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U ✗



# Spam Prediction

We are given a set of labeled email messages

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight ✓

To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U ✗

To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting ✓

Goal is to predict labels of new messages that arrive

To: Jenn Wortman Vaughan  
From: NIPS Committee  
Subject: Paper decision ?

# A Classification Problem

First we need a way to represent the data...

“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

# A Classification Problem

First we need a way to represent the data...

“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

“feature vector”

“label”

# A Classification Problem

First we need a way to represent the data...

“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

Then we need a reasonable set of prediction rules...

- Disjunctions (spam if not known or not “260”)
- Thresholds (spam if “Jenn”+“260”+known < 2)

# A Classification Problem

First we need a way to represent the data...

“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

Then we need a reasonable set of prediction rules...

- Disjunctions (spam if not known or not “260”)
- **Thresholds** (spam if “Jenn”+“260”+known < 2)

“concept class” or “function class” or “hypothesis class”

# A Classification Problem

First we need a way to represent the data...

“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

Then we need a reasonable set of prediction rules...

- Disjunctions (spam if not known or not “260”)
- Thresholds (spam if “Jenn”+“260”+known < 2)

“prediction rule” or “hypothesis” or “concept”

# A Classification Problem

First we need a way to represent the data...

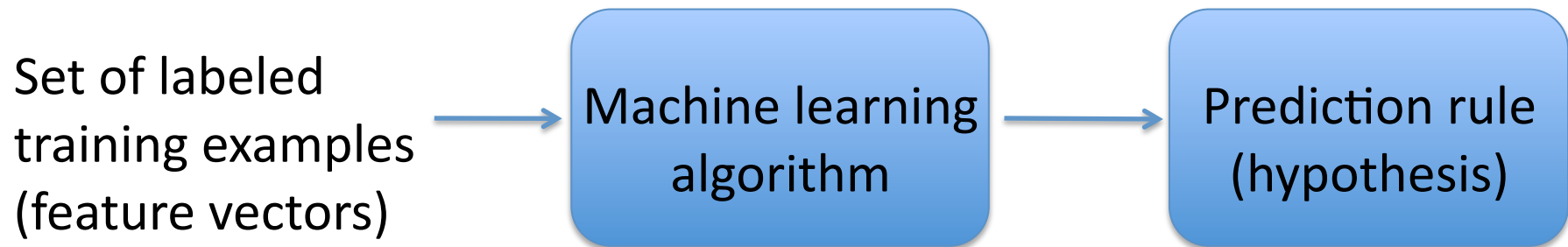
“Jenn”	“260”	“Viagra”	Known Sender	Spelling Bad	Spam?
1	1	0	0	1	0
1	0	0	1	0	0
0	0	1	0	0	1
0	0	0	0	1	1
0	1	0	1	0	0

Then we need a reasonable set of prediction rules...

- Disjunctions (spam if not known or not “260”)
- Thresholds (spam if “Jenn”+“260”+known < 2)

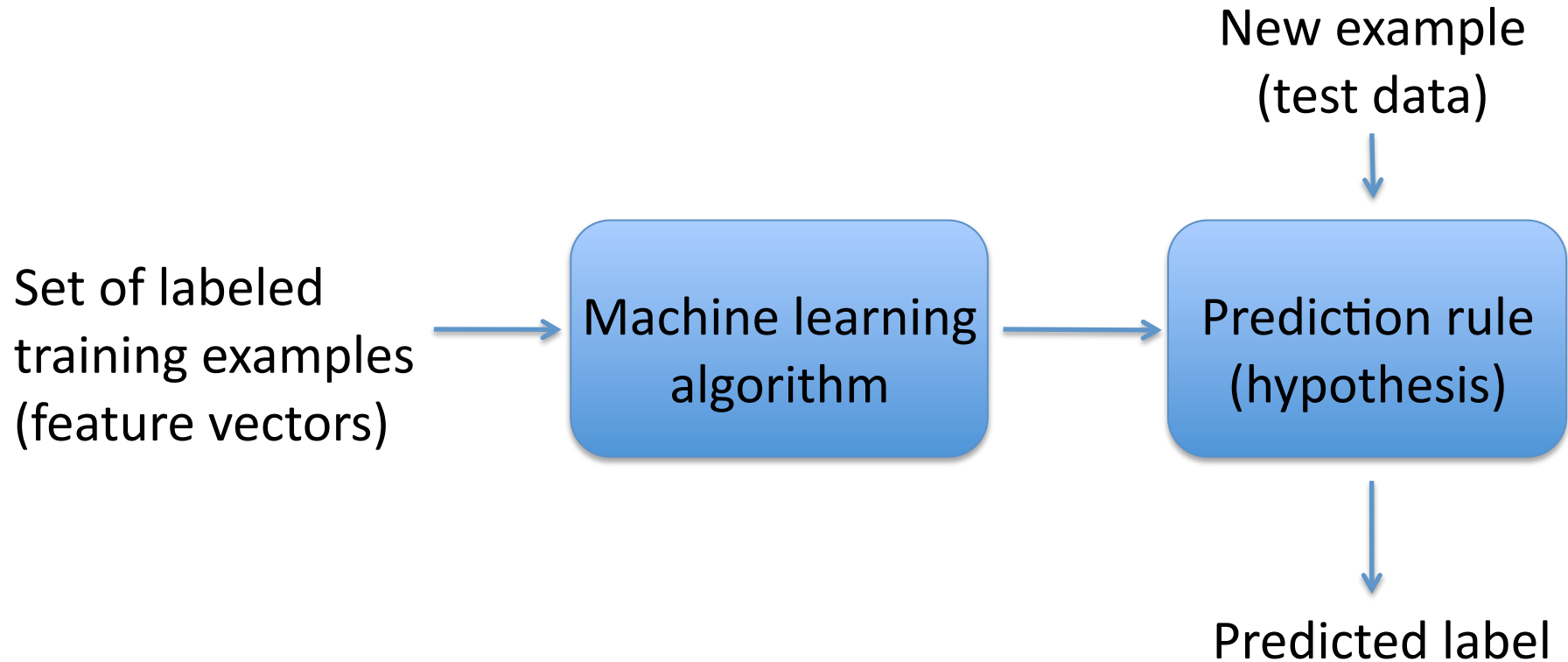
Finally, we need an algorithm...

# Typical Classification Problem





# Typical Classification Problem



# Batch Versus Online Learning

What if there are no clear training and test sets?

# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight

# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting

# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting



# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting



To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U

# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting



To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U





# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting



To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U



# Batch Versus Online Learning

What if there are no clear training and test sets?

To: Jenn Wortman Vaughan  
From: Jeff Vaughan  
Subject: Plans for tonight



To: Jenn Wortman Vaughan  
From: Jens Palsberg  
Subject: Meeting



To: Jenn Wortman Vaughan  
From: Bob Smith  
Subject: V14GR4 4 U



The goal is now to update the prediction rule over time while making as few mistakes as possible

# Other Learning Settings

- Unsupervised learning (clustering)
- Semi-supervised learning
- Active learning
- Reinforcement learning

What is learning *theory*?

# What is learning *theory*?

The goal of learning theory is to develop and analyze formal models that help us understand

... what concepts we can hope to learn efficiently,  
and how much data is necessary to learn them

# What is learning *theory*?

The goal of learning theory is to develop and analyze formal models that help us understand

- ... what concepts we can hope to learn efficiently, and how much data is necessary to learn them
- ... what types of guarantees we might hope to achieve (error bounds, complexity bounds)

# What is learning *theory*?

The goal of learning theory is to develop and analyze formal models that help us understand

- ... what concepts we can hope to learn efficiently, and how much data is necessary to learn them
- ... what types of guarantees we might hope to achieve (error bounds, complexity bounds)
- ... why particular algorithms may or may not perform well under various conditions

# What is learning *theory*?

The goal of learning theory is to develop and analyze formal models that help us understand

... what concepts we can hope to learn efficiently, and how much data is necessary to learn them

... what types of guarantees we might hope to achieve (error bounds, complexity bounds)

... why particular algorithms may or may not perform well under various conditions

This generates intuition useful for algorithm design



# Questions We Ask

- What are the intrinsic properties of a learning problem that impact the amount of data we need?

# Questions We Ask

- What are the intrinsic properties of a learning problem that impact the amount of data we need?
- How much prior information or domain knowledge do we need to learn effectively?

# Questions We Ask

- What are the intrinsic properties of a learning problem that impact the amount of data we need?
- How much prior information or domain knowledge do we need to learn effectively?
- Are simpler hypotheses always better? Why?

# Questions We Ask

- What are the intrinsic properties of a learning problem that impact the amount of data we need?
- How much prior information or domain knowledge do we need to learn effectively?
- Are simpler hypotheses always better? Why?
- How should we trade off the standard notions of efficiency (time, space) with data efficiency?

# Course Overview

1. Classification and the “probably approximately correct” (PAC) model of learning

# Course Overview

1. Classification and the “probably approximately correct” (PAC) model of learning
2. Online learning in adversarial settings, including the “expert advice” framework

# Course Overview

1. Classification and the “probably approximately correct” (PAC) model of learning
2. Online learning in adversarial settings, including the “expert advice” framework
3. Some of learning theory’s success stories, including boosting and support vector machines

# Course Overview

1. Classification and the “probably approximately correct” (PAC) model of learning
2. Online learning in adversarial settings, including the “expert advice” framework
3. Some of learning theory’s success stories, including boosting and support vector machines
4. Recent research topics (time permitting)



# Things We Will NOT Cover

- Implementation tricks
  - Feature design
  - Particular application domains (natural language, vision, robotics, search, etc.)
  - Commercial uses of machine learning
- ... but you are welcome to explore some of these topics as one part of your final project

# Who should take this class?

- Students with a background in machine learning who want to know more about the theoretical foundations
- Students with a background in CS theory who want to know more about learning
- Prerequisites: comfort with probability theory, comfort reading & writing mathematical proofs

# Course Logistics

# Reading Material

There is **no required textbook** for this class

Lecture notes and links to supplementary reading material will be posted regularly at:

<http://www.cs.ucla.edu/~jenn/courses/F11.html>

Check this website often!!

# Breakdown of Grades

Four homework assignments (60%)

- Mostly analysis and proofs
- No coding required

Final Project (40%)

- In-class presentation and written report

# Academic Honesty Policy

Collaboration is **strongly encouraged**, but...

- Each student must write down his or her own solutions independently in his or her own words.
- Each student must submit a list of anyone with whom the assignment was discussed.
- All sources (internet included) must be credited.
- Solution sets from this course or any other course may not be used under any circumstances.

# PTEs

If you would like a PTE for this class, please

- Come to class today and Wednesday
- Send me an email **after Wednesday's class** telling me a little about your background

# Logistical Loose Ends

- This course counts toward AI majors/minors
- Auditors are welcome **as long as there is space** – please reserve seats for students who are enrolled
- All of this info and more is on the course website



# Models of Learning

# Models of Learning

- A learning model must specify several things
  - What are we trying to learn?
  - What kind of data is available?
  - How is the data presented to the learner?
  - What type of feedback does the learner receive?
  - What is the goal of the learning process?

# Models of Learning

- A learning model must specify several things
  - What are we trying to learn?
  - What kind of data is available?
  - How is the data presented to the learner?
  - What type of feedback does the learner receive?
  - What is the goal of the learning process?
- To provide valuable insight, a learning model must be **robust to minor variations** in its definition

# The Consistency Model

- **Definition:** We say that algorithm  $A$  learns concept class  $C$  in the consistency model if given a set of labeled examples  $S$ ,  $A$  produces a concept  $c \in C$  consistent with  $S$  if one exists and states that none exists otherwise.

# The Consistency Model

- **Definition:** We say that algorithm  $A$  learns concept class  $C$  in the consistency model if given a set of labeled examples  $S$ ,  $A$  produces a concept  $c \in C$  consistent with  $S$  if one exists and states that none exists otherwise.
- **Definition:** We say that a class  $C$  is **efficiently learnable** in the consistency model if there exists an efficient algorithm  $A$  that learns  $C$ .

# Example: Monotone Conjunctions

Guitar	Fast beat	Male singer	Acoustic	New	Liked
1	0	0	1	1	1
1	1	1	0	0	0
0	1	1	0	1	0
1	0	1	1	0	1
1	0	0	0	1	0

# Example: Monotone Conjunctions

Guitar	Fast beat	Male singer	Acoustic	New	Liked
1	0	0	1	1	1
1	1	1	0	0	0
0	1	1	0	1	0
1	0	1	1	0	1
1	0	0	0	1	0

- Find all of the variables that are true in every positive example.
- Let  $c$  be the conjunction of these variables.
- Output  $c$  if it is consistent with all negative examples; otherwise, output none.

# Example: DNFs

Guitar	Fast beat	Male singer	Acoustic	New	Liked
1	0	0	1	1	1
1	1	1	0	0	0
0	1	1	0	1	0
1	0	1	1	0	1
1	0	0	0	1	0

DNF = the set of all disjunctions of conjunctions



# Example: DNFs

Guitar	Fast beat	Male singer	Acoustic	New	Liked
1	0	0	1	1	1
1	1	1	0	0	0
0	1	1	0	1	0
1	0	1	1	0	1
1	0	0	0	1	0

DNF = the set of all disjunctions of conjunctions

Trivial to learn in the consistency model!

What is wrong with this model?