CSE 158 – Lecture 1.5

Web Mining and Recommender Systems

Supervised learning – Regression

What is supervised learning?

Supervised learning is the process of trying to infer from **labeled data** the underlying function that produced the labels associated with the data

What is supervised learning?

Given **labeled training data** of the form

$$\{(\text{data}_1, \text{label}_1), \ldots, (\text{data}_n, \text{label}_n)\}$$

Infer the function

$$f(\text{data}) \stackrel{?}{\rightarrow} \text{labels}$$

Example

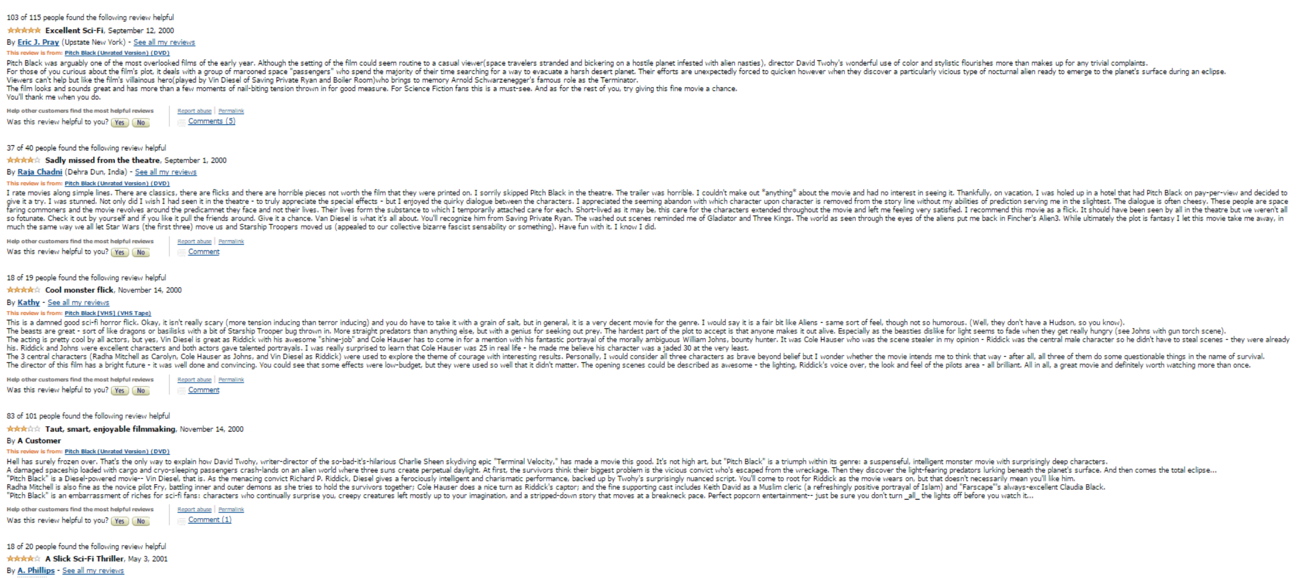
Suppose we want to build a movie recommender

e.g. which of these films will I rate highest?

Example

Q: What are the **labels?**

A: ratings that others have given to each movie, and that I have given to other movies



103 of 115 people found the following review helpful

Excellent Sci-Fi, September 12, 2000

By Eric J. Pray (Upstate New York) - See all my reviews

This review is from: Pitch Black (Unrated Version) (DVD)

Pitch Black was arguably one of the most overlooked films of the early year. Although the setting of the film could seem routine to a casual viewer(space travelers stranded and bickering on a hostile planet infested with alien nasties), director David Twohy's wonderful use of color and stylistic flourishes more than makes up for any trivial complaints.
For those of you curious about the film's plot, it deals with a group of marooned space "passengers" who spend the majority of their time searching for a way to evacuate a harsh desert planet. Their efforts are unexpectedly forced to quicken however when they discover a particularly vicious type of nocturnal alien ready to emerge to the planet's surface during an eclipse.
Viewers can't help but like the film's villainous hero(played by Vin Diesel of Saving Private Ryan and Boiler Room)who brings to memory Arnold Schwarzenegger's famous role as the Terminator.
The film looks and sounds great and has more than a few moments of nail-biting tension thrown in for good measure. For Science Fiction fans this is a must-see. And as for the rest of you, try giving this fine movie a chance.
You'll thank me when you do.

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comments (5)

37 of 40 people found the following review helpful

Sadly missed from the theatre, September 1, 2000

By Raja Chadni (Dehra Dun, India) - See all my reviews

This review is from: Pitch Black (Unrated Version) (DVD)

I rate movies along simple lines. There are classics, there are flicks and there are horrible pieces not worth the film that they were printed on. I sorrily skipped Pitch Black in the theatre. The trailer was horrible. I couldn't make out *anything* about the movie and had no interest in seeing it. Thankfully, on vacation, I was holed up in a hotel that had Pitch Black on pay-per-view and decided to give it a try. I was stunned. Not only did I wish I had seen it in the theatre - to truly appreciate the special effects - but I enjoyed the quirky dialogue between the characters. I appreciated the seeming abandon with which character upon character is removed from the story line without my abilities of prediction serving me in the slightest. The dialogue is often cheesy. These people are space faring commoners and the movie revolves around the predicamnet they face and not their lives. Their lives form the substance to which I temporarily attached care for each. Short-lived as it may be, this care for the characters extended throughout the movie and left me feeling very satisfied. I recommend this movie as a flick. It should have been seen by all in the theatre but we weren't all so fotunate. Check it out by yourself and if you like it pull the friends around. Give it a chance. Van Diesel is what it's all about. You'll recognize him from Saving Private Ryan. The washed out scenes reminded me of Gladiator and Three Kings. The world as seen through the eyes of the aliens put me back in Fincher's Alien3. While ultimately the plot is fantasy I let this movie take me away, in much the same way we all let Star Wars (the first three) move us and Starship Troopers moved us (appealed to our collective bizarre fascist sensability or something). Have fun with it. I know I did.

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comment

18 of 19 people found the following review helpful

Cool monster flick, November 14, 2000

By Kathy - See all my reviews

This review is from: Pitch Black [VHS] (VHS Tape)

This is a damned good sci-fi horror flick. Okay, it isn't really scary (more tension inducing than terror inducing) and you do have to take it with a grain of salt, but in general, it is a very decent movie for the genre. I would say it is a fair bit like Aliens - same sort of feel, though not so humorous. (Well, they don't have a Hudson, so you know).
The beasts are great - sort of like dragons or basilisks with a bit of Starship Trooper bug thrown in. More straight predators than anything else, but with a genius for seeking out prey. The hardest part of the plot to accept is that anyone makes it out alive. Especially as the beasties dislike for light seems to fade when they get really hungry (see Johns with gun torch scene).
The acting is pretty cool by all actors, but yes, Vin Diesel is great as Riddick with his awesome "shine-job" and Cole Hauser has to come in for a mention with his fantastic portrayal of the morally ambiguous William Johns, bounty hunter. It was Cole Hauser who was the scene stealer in my opinion - Riddick was the central male character so he didn't have to steal scenes - they were already his. Riddick and Johns were excellent characters and both actors gave talented portrayals. I was really surprised to learn that Cole Hauser was 25 in real life - he made me believe his character was a jaded 30 at the very least.
The 3 central characters (Radha Mitchell as Carolyn, Cole Hauser as Johns, and Vin Diesel as Riddick) were used to explore the theme of courage with interesting results. Personally, I would consider all three characters as brave beyond belief but I wonder whether the movie intends me to think that way - after all, all three of them do some questionable things in the name of survival.
The director of this film has a bright future - it was well done and convincing. You could see that some effects were low-budget, but they were used so well that it didn't matter. The opening scenes could be described as awesome - the lighting, Riddick's voice over, the look and feel of the pilots area - all brilliant. All in all, a great movie and definitely worth watching more than once.

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comment

83 of 101 people found the following review helpful

Taut, smart, enjoyable filmmaking, November 14, 2000

By **A Customer**

This review is from: Pitch Black (Unrated Version) (DVD)

Hell has surely frozen over. That's the only way to explain how David Twohy, writer-director of the so-bad-it's-hilarious Charlie Sheen skydiving epic "Terminal Velocity," has made a movie this good. It's not high art, but "Pitch Black" is a triumph within its genre: a suspenseful, intelligent monster movie with surprisingly deep characters.
A damaged spaceship loaded with cargo and cryo-sleeping passengers crash-lands on an alien world where three suns create perpetual daylight. At first, the survivors think their biggest problem is the vicious convict who's escaped from the wreckage. Then they discover the light-fearing predators lurking beneath the planet's surface. And then comes the total eclipse...
"Pitch Black" is a Diesel-powered movie-- Vin Diesel, that is. As the menacing convict Richard P. Riddick, Diesel gives a ferociously intelligent and charismatic performance, backed up by Twohy's surprisingly nuanced script. You'll come to root for Riddick as the movie wears on, but that doesn't necessarily mean you'll like him.
Radha Mitchell is also fine as the novice pilot Fry, battling inner and outer demons as she tries to hold the survivors together; Cole Hauser does a nice turn as Riddick's captor; and the fine supporting cast includes Keith David as a Muslim cleric (a refreshingly positive portrayal of Islam) and "Farscape"'s always-excellent Claudia Black.
"Pitch Black" is an embarrassment of riches for sci-fi fans: characters who continually surprise you, creepy creatures left mostly up to your imagination, and a stripped-down story that moves at a breakneck pace. Perfect popcorn entertainment-- just be sure you don't turn _all_ the lights off before you watch it...

Help other customers find the most helpful reviews | Report abuse | Permalink
Was this review helpful to you? Yes No | Comment (1)

18 of 20 people found the following review helpful

A Slick Sci-Fi Thriller, May 3, 2001

By A. Phillips - See all my reviews

REAL NAME

Example

Q: What is the **data?**

A: **features** about the movie and the users who evaluated it

Movie features: genre, actors, rating, length, etc.

Product Details

Genres	Science Fiction, Action, Horror
Director	David Twohy
Starring	Vin Diesel, Radha Mitchell
Supporting actors	Cole Hauser, Keith David, Lewis Fitz-Gerald, Claudia Black, Rhiana Gr Angela Moore, Peter Chiang, Ken Twohy
Studio	NBC Universal
MPAA rating	R (Restricted)
Captions and subtitles	English Details
Rental rights	24 hour viewing period. Details
Purchase rights	Stream instantly and download to 2 locations Details
Format	Amazon Instant Video (streaming online video and digital download)

User features: age, gender, location, etc.

A. Phillips

Reviewer ranking: #17,230,554

90% helpful

votes received on reviews (151 of 167)

ABOUT ME

Enjoy the reviews...

ACTIVITIES

Reviews (16)

Public Wish List (2)

Listmania Lists (2)

Tagged Items (1)

Example

Movie recommendation:

$$f(\text{data}) \overset{?}{\rightarrow} \text{labels}$$

=

$$f(\text{user features}, \text{movie features}) \overset{?}{\rightarrow} \text{star rating}$$

Solution 1

Design a system based on **prior knowledge**, e.g.

```
def prediction(user, movie):
    if (user['age'] <= 14):
        if (movie['mpaa_rating']) == "G"):
            return 5.0
        else:
            return 1.0
    else if (user['age'] <= 18):
        if (movie['mpaa_rating']) == "PG"):
            return 5.0
….. Etc.
```

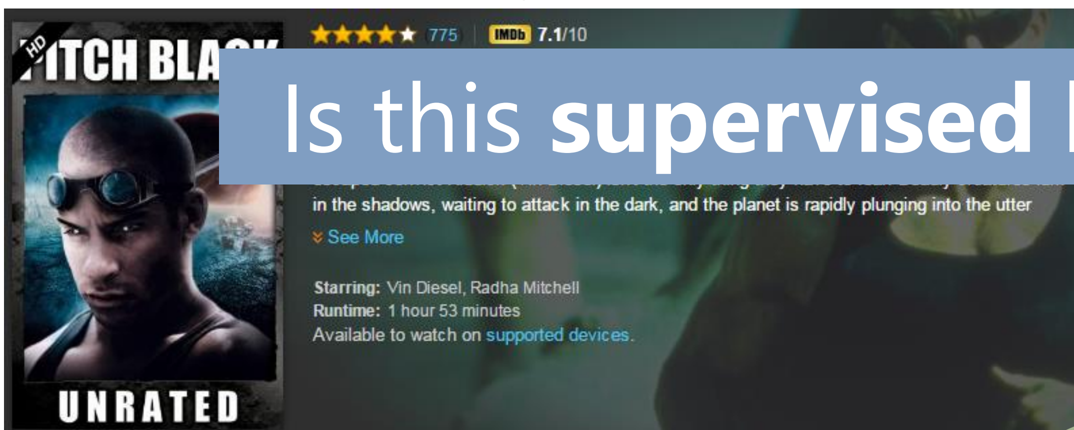
Is this **supervised learning?**

Solution 2

Identify words that I frequently mention in my social media posts, and recommend movies whose plot synopses use **similar** types of language

Social media posts

Plot synopsis



Is this **supervised learning?**

argmax similarity(synopsis, post)

Solution 3

Identify which attributes (e.g. actors, genres) are associated with positive ratings. Recommend movies that exhibit those attributes.

Is this **supervised learning?**

Solution 1

(design a system based on prior knowledge)

Disadvantages:

- Depends on possibly false **assumptions** about how users relate to items
- Cannot adapt to new data/information

Advantages:

- Requires no data!

Solution 2

(identify similarity between wall posts and synopses)

Disadvantages:

- Depends on possibly false **assumptions** about how users relate to items
- May not be adaptable to new settings

Advantages:

- Requires data, but does not require **labeled** data

Solution 3

(identify attributes that are associated with positive ratings)

Disadvantages:

- Requires a (possibly large) dataset of movies with labeled ratings

Advantages:

- Directly optimizes a measure we care about (predicting ratings)
- Easy to adapt to new settings and data

Supervised versus unsupervised learning

Learning approaches attempt to **model data** in order to solve a problem

Unsupervised learning approaches find patterns/relationships/structure in data, but **are not** optimized to solve a particular predictive task 2

Supervised learning aims to directly model the relationship between input and output variables, so that the output variables can be predicted accurately given the input 3

Not learning 1

Regression

Regression is one of the simplest supervised learning approaches to learn relationships between input variables (features) and output variables (predictions)

Linear regression

Linear regression assumes a predictor of the form

feat of movie1

movie2

$$X\theta = y \qquad \begin{bmatrix}5\\3\\2\\4\end{bmatrix}\ \vdots$$

matrix of features (data)

unknowns (which features are relevant)

vector of outputs (labels)

(or $Ax = b$ if you prefer)

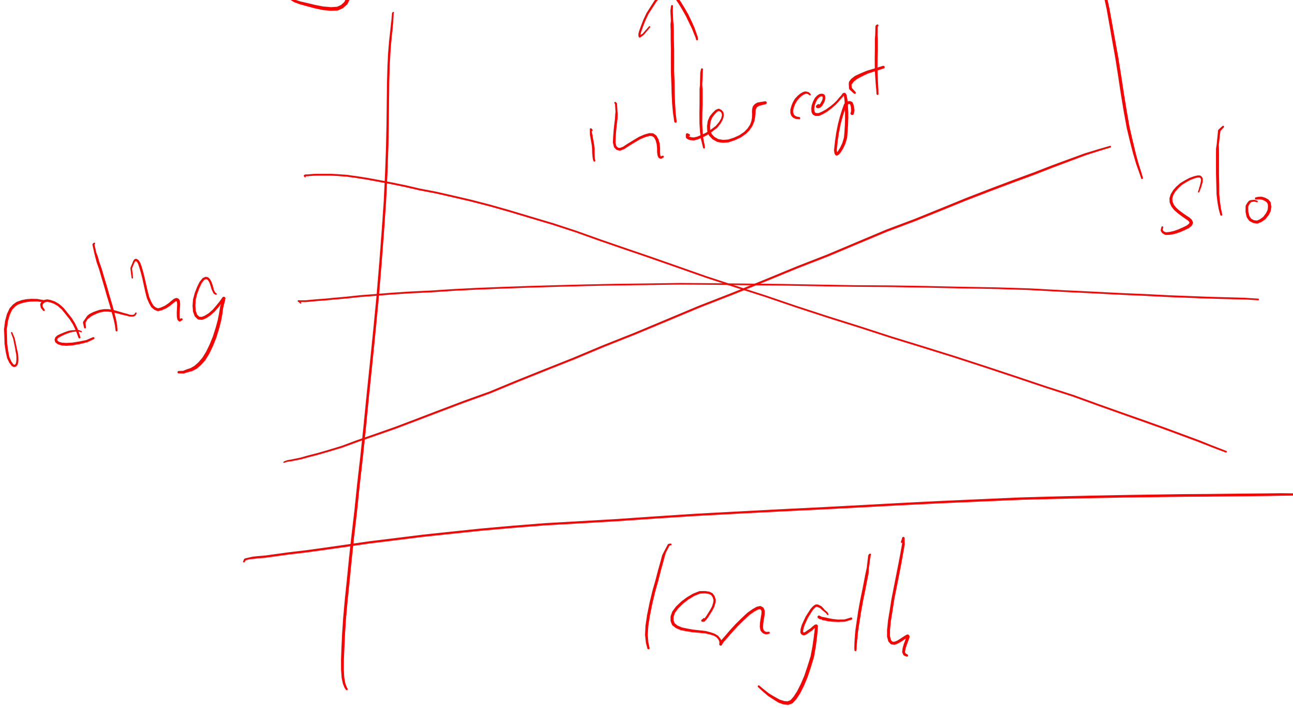
$$\text{rating} = \langle \theta, \text{features} \rangle$$

$$y_i = \langle \theta, x_i \rangle = \langle [\theta_0, \theta_1], [1, \text{length}] \rangle$$

$$\text{rating} = \theta_0 + \theta_1 \, \text{length}$$

intercept

slope



Linear regression

Linear regression assumes a predictor of the form

$$\begin{bmatrix} 1 & \text{length}_1 \\ 1 & \text{length}_2 \\ \vdots & \vdots \\ 1 & \text{length}_N \end{bmatrix} X\theta = y \quad \begin{bmatrix} \text{rating}_1 \\ \text{rating}_2 \\ \vdots \end{bmatrix}$$

Q: Solve for theta

A:

$$\theta = (X^T X)^{-1} X^T y$$

Example 1

How do preferences toward certain beers vary with age?

Example 1

Beeradvocate

Beers:

Displayed for educational use only; do not reuse.

BA SCORE	THE BROS	
100	95	Ratings: 9,587
world-class	world-class	Reviews: 2,537
		rAvg: 4.59
		pDev: 9.59%
		Wants: 2,109
9,587 Ratings	(view ratings)	Gots: 4,563 | FT: 472

Brewed by:
Goose Island Beer Co.
Illinois, United States

Style | ABV
American Double / Imperial Stout | 13.80% ABV

Availability: Winter

Notes/Commercial Description:
60 IBU

(Beer added by: drewbage on 06-26-2003)

Ratings/reviews:

4.35/5 rDev -5.2%
look: 4 | smell: 4.25 | taste: 4.5 | feel: 4.25 | overall: 4.25

Serving: 355 mL bottle poured into a 9 oz Libbey Embassy snifter ("bottled on: 08AUG14 1109").

Appearance: Deep, dark near-black brown. Hazy, light brown fringe of foam and limited lacing; no head.

Smell: Roasted malt, vanilla, and some warming alcohol.

Taste: Roasted malts, cocoa, burnt caramel, molasses, vanilla and dark fruit. Bourbon barrel is hinted at but never takes over.

Mouthfeel: Medium to full body and light carbonation with a very lush, silky smooth feel.

Overall: Not as complex or intense as some newer barrel-aged stouts, but so smooth and balanced with all the elements tightly integrated.

HipCzech, Yesterday at 05:38 AM

User profiles:

HipCzech
Aficionado
Male, from **Texas**
Profile Page

Member Since:	Jul 12, 2014	HipCzech was last seen:
Points:	175	Today at 12:19 AM
Beers:	108	
Places:	6	
Posts:	0	
Likes Received:	0	
Trading:	0% | 0	

Example 1

50,000 reviews are available on
http://jmcauley.ucsd.edu/cse158/data/beer/beer_50000.json
(see course webpage)

See also – non-alcoholic beers:
http://jmcauley.ucsd.edu/cse158/data/beer/non-alcoholic-beer.json

Example 1

Real-valued features

How do preferences toward certain beers vary with age?
How about **ABV**?

$$\text{rating} = \theta_0 + \theta_1 [\text{user age}]$$

(code for all examples is on http://jmcauley.ucsd.edu/cse158/code/week1.py)

Example 1

Real-valued features

What is the interpretation of:

$$\theta = (3.4, 10e^{-7})$$

(code for all examples is on http://jmcauley.ucsd.edu/cse158/code/week1.py)

Example 2

Categorical features

How do beer preferences vary as a function of **gender**?

(code for all examples is on http://jmcauley.ucsd.edu/cse158/code/week1.py)

Example 3

Random features

What happens as we add more and more **random** features?

(code for all examples is on http://jmcauley.ucsd.edu/cse158/code/week1.py)

Exercise

How would you build a feature to represent the **month**, and the impact it has on people's rating behavior?