

CSE 446: Machine Learning

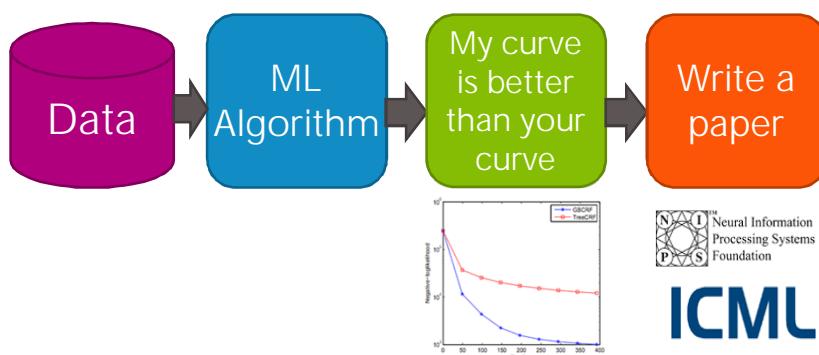
Welcome

Emily Fox
University of Washington
January 4, 2017

©2017 Emily Fox

Machine learning is
changing the world

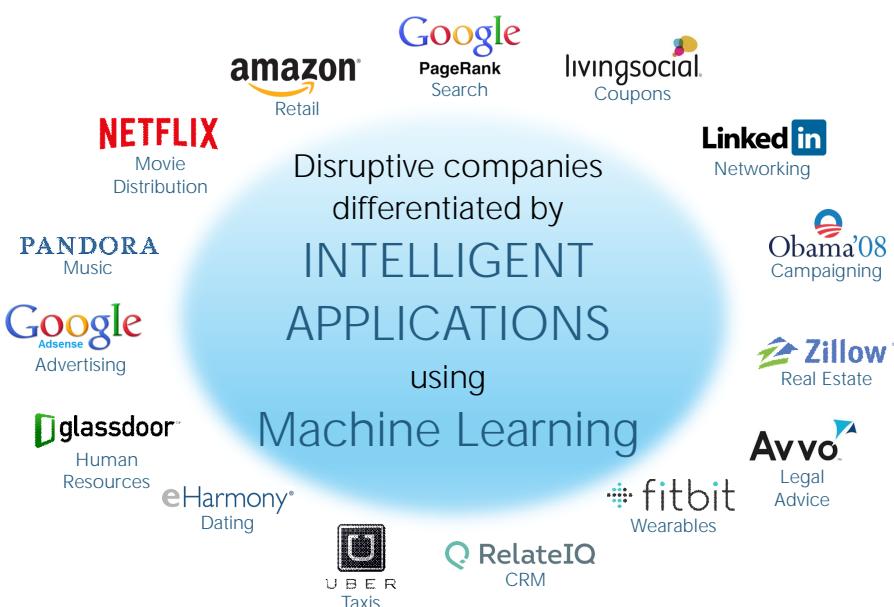
Old view of ML



©2017 Emily Fox

CSE 446: Machine Learning

Disruptive companies differentiated by
INTELLIGENT APPLICATIONS
 using
Machine Learning



What is machine learning?

Generically...

Study of algorithms that
improve their **performance**
at some **task**
with **experience**

The machine learning pipeline



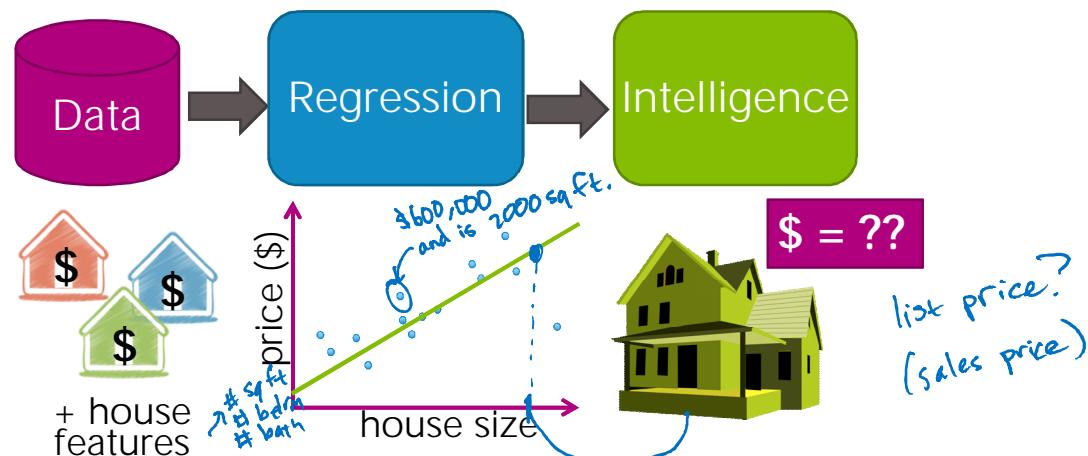
7

©2017 Emily Fox

CSE 446: Machine Learning

Regression

Example: Predicting house prices



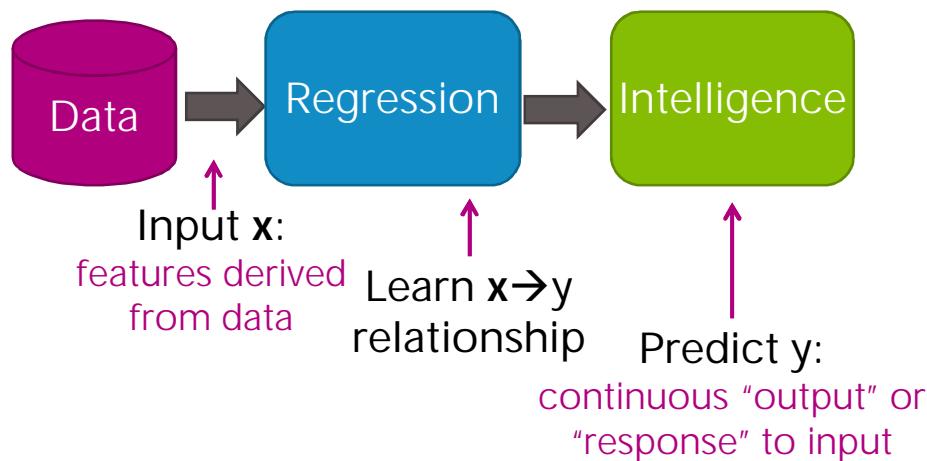
8

©2017 Emily Fox

CSE 446: Machine Learning

What is regression?

From features to predictions

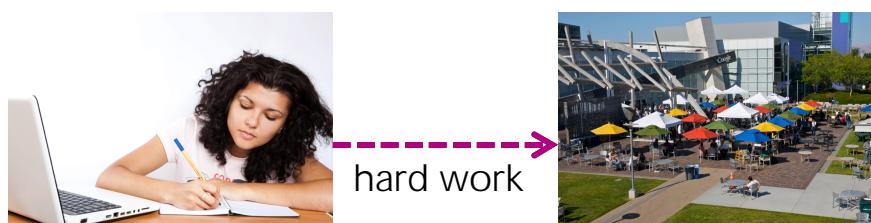


9

©2017 Emily Fox

CSF 446: Machine Learning

Salary after CSE 446



- How much will your salary be? ($y = \text{ $$}$)
- Depends on $x = \text{performance in course, quality of project, # of discussion board responses, ...}$

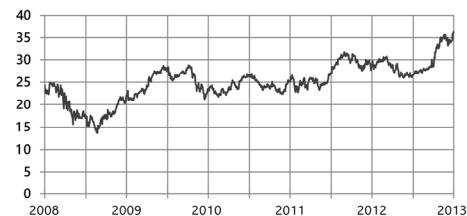
10

©2017 Emily Fox

CSF 446: Machine Learning

Stock prediction

- Predict the price of a stock (y)
- Depends on x =
 - Recent history of stock price
 - News events
 - Related commodities



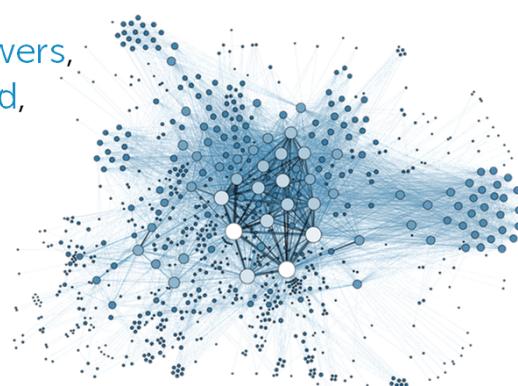
11

©2017 Emily Fox

CSE 446: Machine Learning

Tweet popularity

- How many people will retweet your tweet? (y)
- Depends on x = # followers,
 # of followers of followers,
 features of text tweeted,
 popularity of hashtag,
 # of past retweets,...

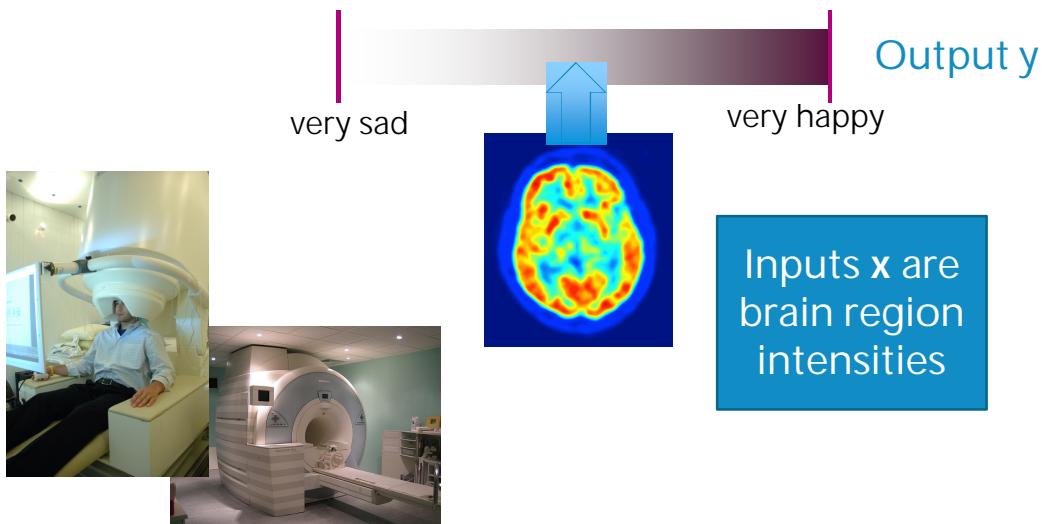


12

©2017 Emily Fox

CSE 446: Machine Learning

Reading your mind



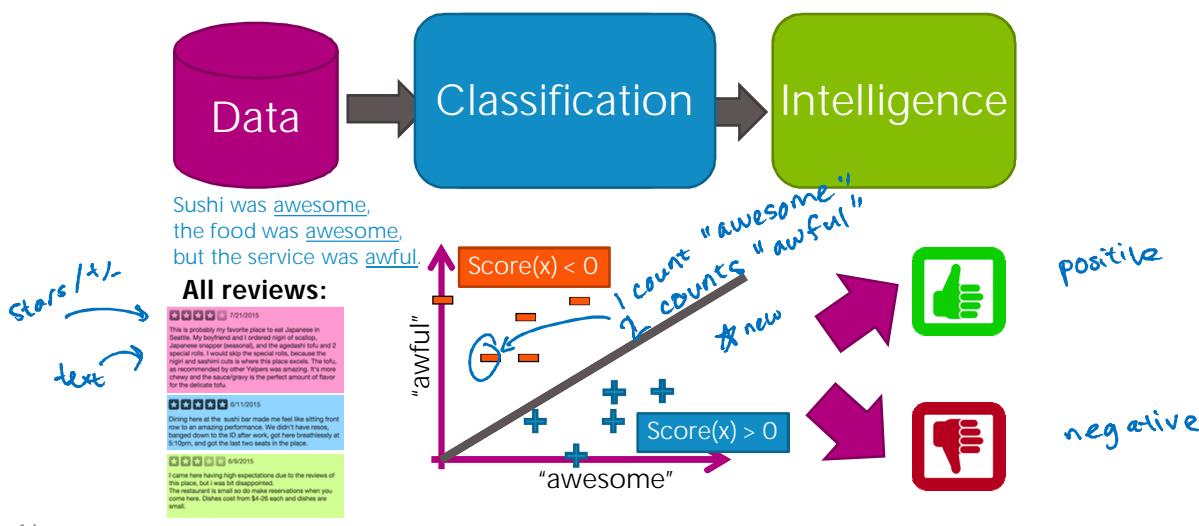
13

©2017 Emily Fox

CSF 446: Machine Learning

Classification

Example: Sentiment analysis



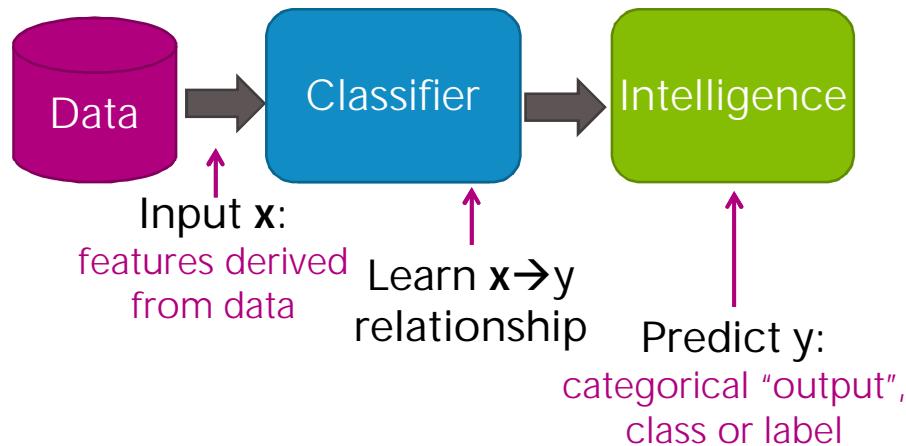
14

©2017 Emily Fox

CSE 446: Machine Learning

What is classification?

From features to predictions

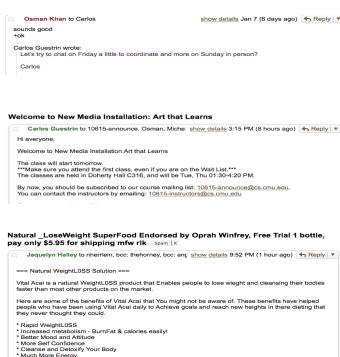


15

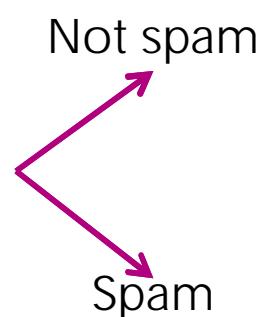
©2017 Emily Fox

CSF 446: Machine Learning

Spam filtering



**Text of email,
sender, IP,...**



Input: x

Output: y

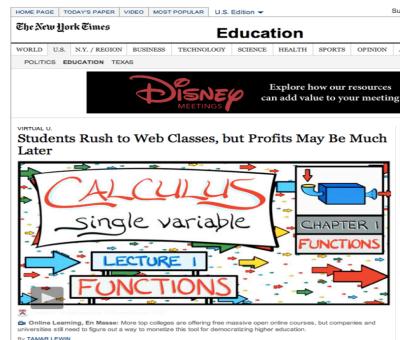
16

©2017 Emily Fox

CSF 446: Machine Learning

Multiclass classifier

Output y has more than 2 categories



Input: x
Webpage

Education
Finance
Technology

Output: y

17

©2017 Emily Fox

CSF 446: Machine Learning

Image classification



Input: x
Image pixels



Soft predictions

Output: y
Predicted object

18

©2017 Emily Fox

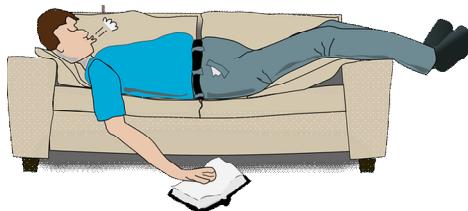
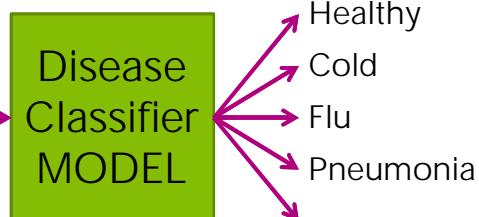
CSF 446: Machine Learning

Personalized medical diagnosis

Input: x



Output: y



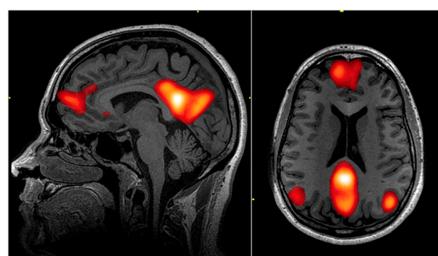
19

©2017 Emily Fox

CSE 446: Machine Learning

Reading your mind

Output y



Inputs x are
brain region
intensities



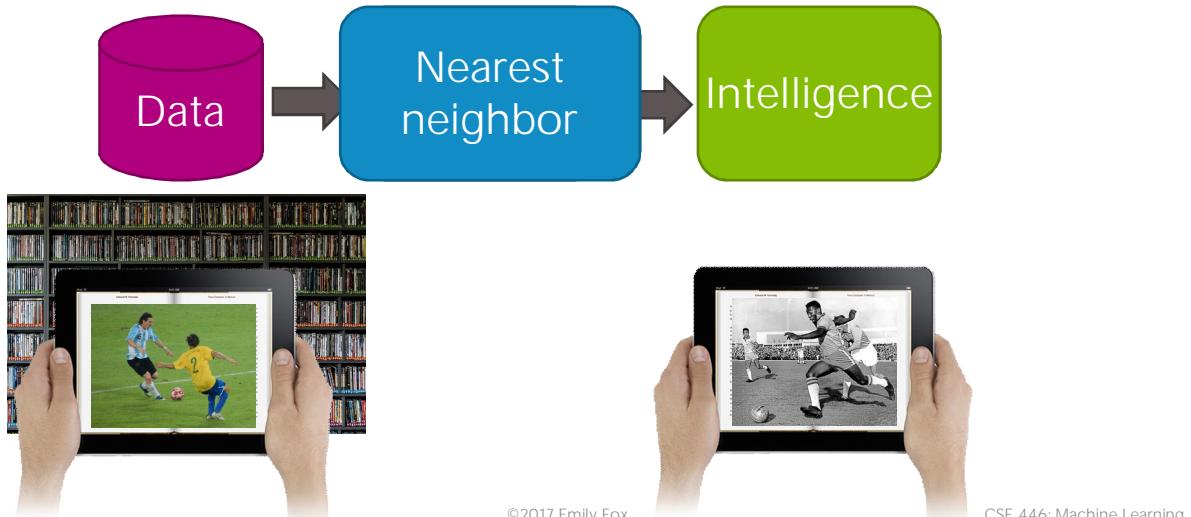
20

©2017 Emily Fox

CSE 446: Machine Learning

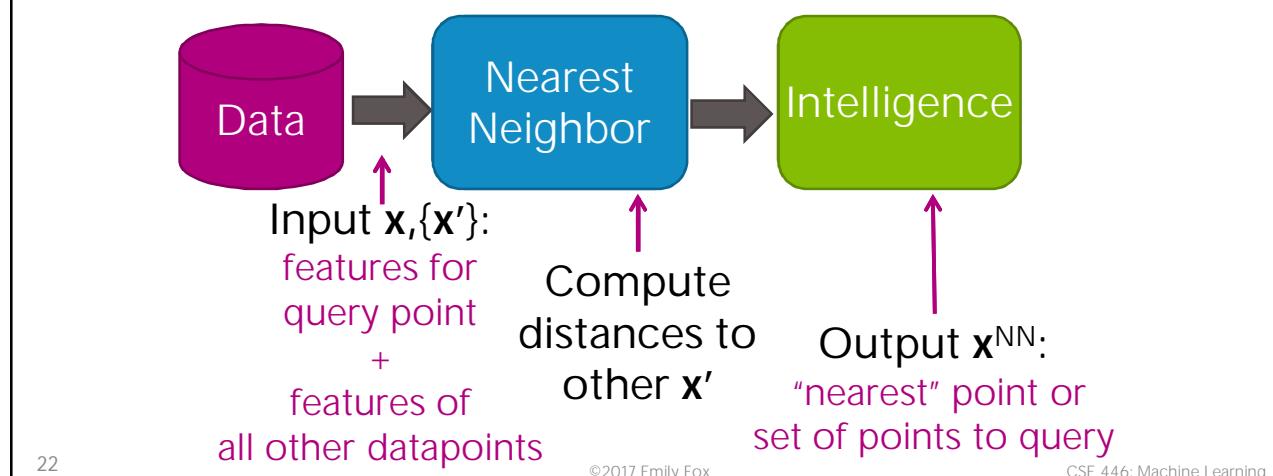
Similarity/finding data

Example: Document retrieval



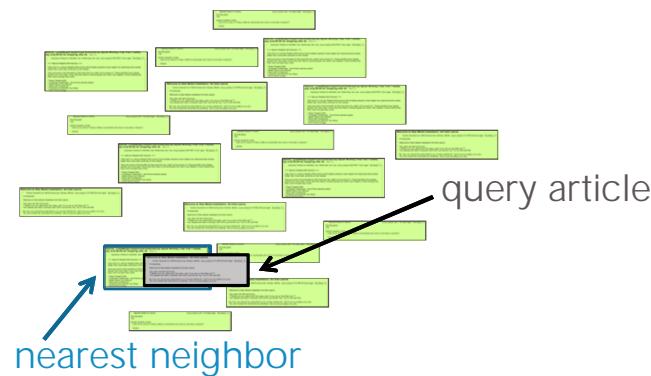
What is retrieval?

Search for related items



Retrieve “nearest neighbor” article

Space of all articles,
organized by similarity of text



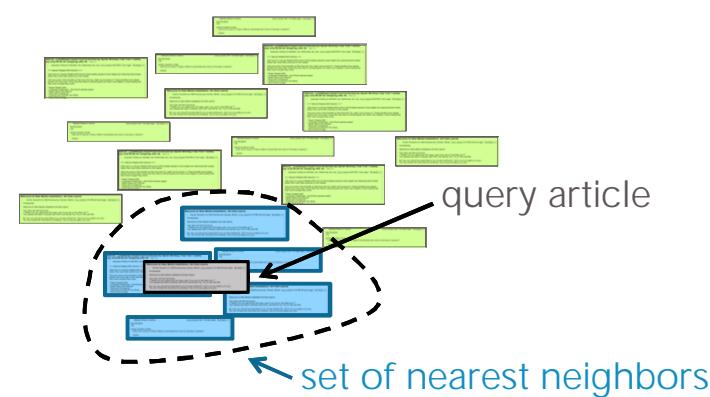
23

©2017 Emily Fox

CSE 446: Machine Learning

Or set of nearest neighbors

Space of all articles,
organized by similarity of text



24

©2017 Emily Fox

CSE 446: Machine Learning

Retrieval applications

Just about everything...

Images



Products



Streaming content:

- Songs
- Movies
- TV shows
- ...

News articles



Social networks

(people you might want to connect with)



25

©2017 Emily Fox

CSF 446: Machine Learning

Clustering

Example: Document structuring for retrieval



SPORTS



WORLD NEWS



ENTERTAINMENT



SCIENCE

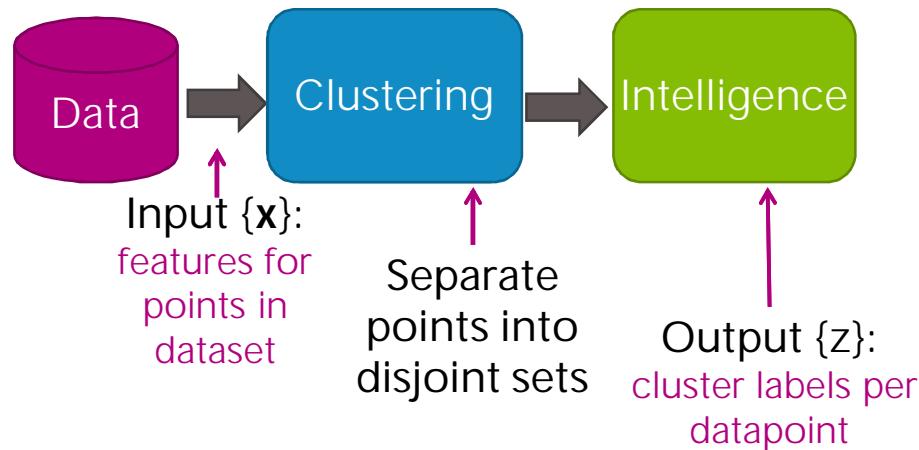
26

©2017 Emily Fox

CSF 446: Machine Learning

What is clustering?

Discover groups of similar inputs



27

©2017 Emily Fox

CSF 446: Machine Learning

Clustering images

For search, group as:

- Ocean
- Pink flower
- Dog
- Sunset
- Clouds
- ...



28

©2017 Emily Fox

CSF 446: Machine Learning

Or users on websites...

Discover groups of users for better targeting of content



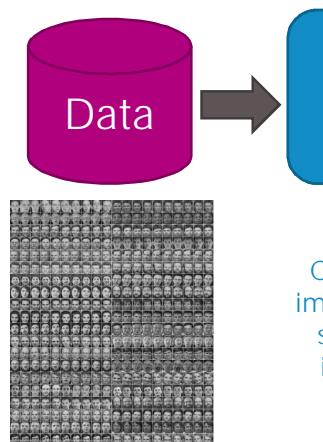
29

©2017 Emily Fox

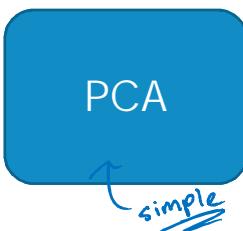
CSE 446: Machine Learning

Embedding

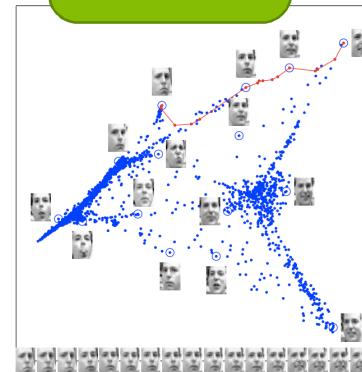
Example: Embedding images to visualize data



Images with thousands or millions of pixels



Can we give each image a coordinate, such that similar images are near each other?



30

©2017 Emily Fox

CSE 446: Machine Learning

Embedding words



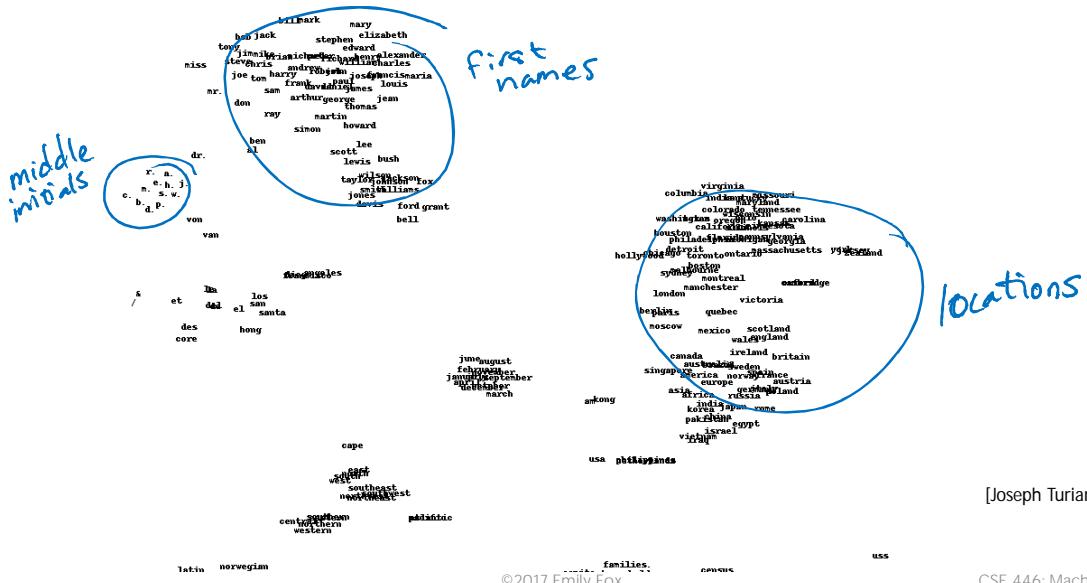
[Joseph Turian]

31

©2017 Emily Fox

CSF 446: Machine Learning

Embedding words (zoom in)



[Joseph Turian]

32

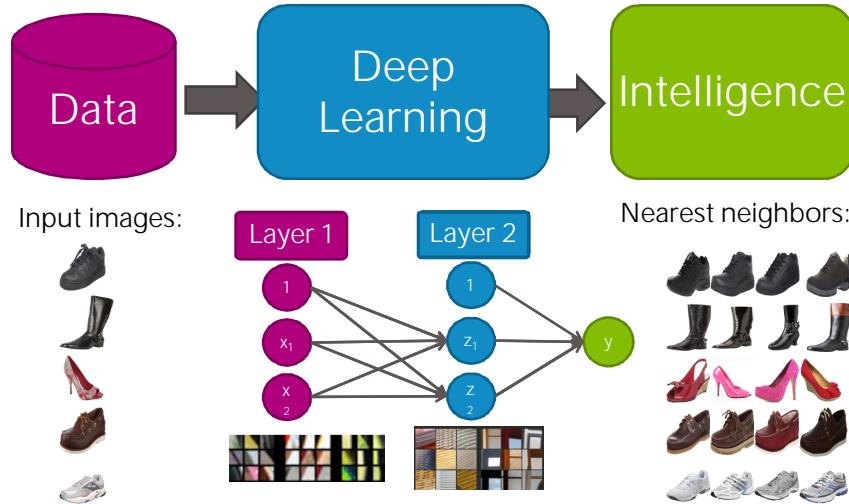
latin norwegian families census

©2017 Emily Fox

CSF 446: Machine Learning

Deep Learning

Example: Visual product recommender



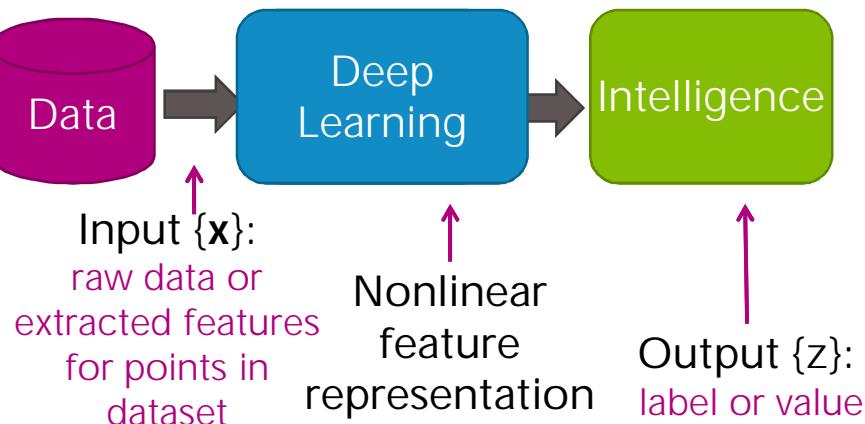
33

©2017 Emily Fox

CSF 446: Machine Learning

What is (supervised) deep learning?

Flexible method for performing classification or regression

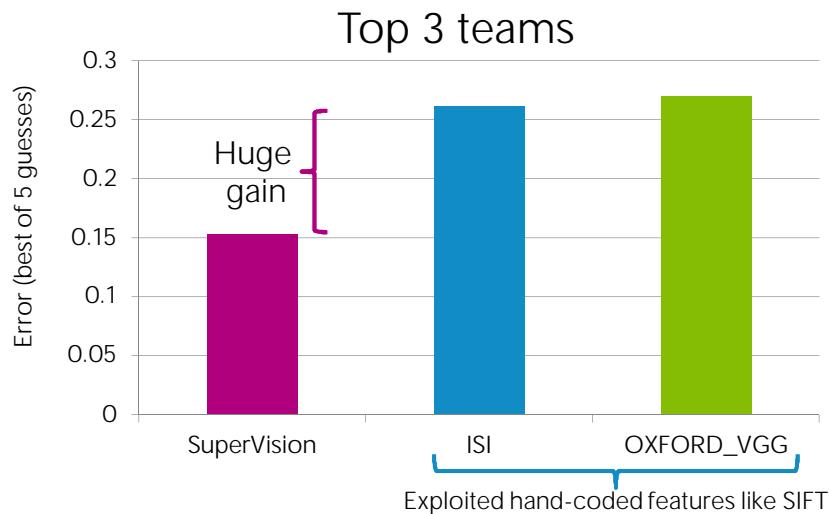


34

©2017 Emily Fox

CSF 446: Machine Learning

ImageNet 2012 competition: 1.2M training images, 1000 categories



©2017 Emily Fox

CSF 446: Machine Learning

Examples of deep learning success stories

- Image classification
- Image segmentation
- Image captioning
- Object detection
- Speech recognition
- Speech synthesis
- Machine translation
- Handwriting recognition
- ...

Other ML topics we won't cover

- Recommender systems
- Reinforcement learning
- Learning theory
- Active learning
- Multi-task and transfer learning
- Spectral methods
- ...

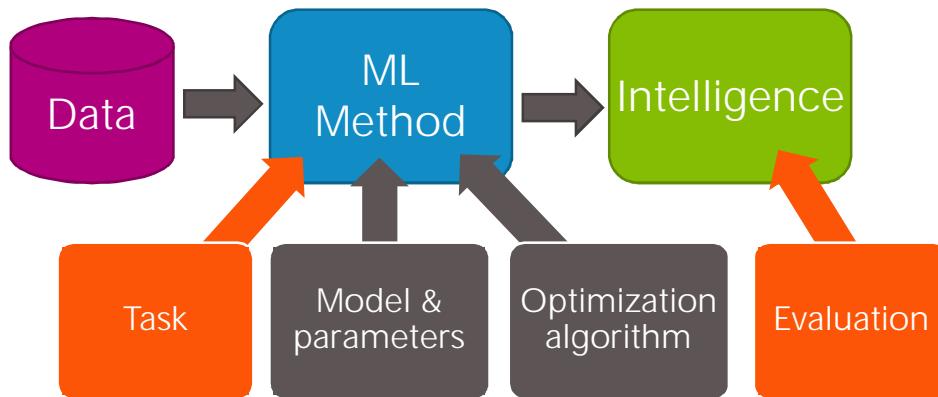
37

©2017 Emily Fox

CSF 446: Machine Learning

Syllabus

Will learn about the ML pipeline...

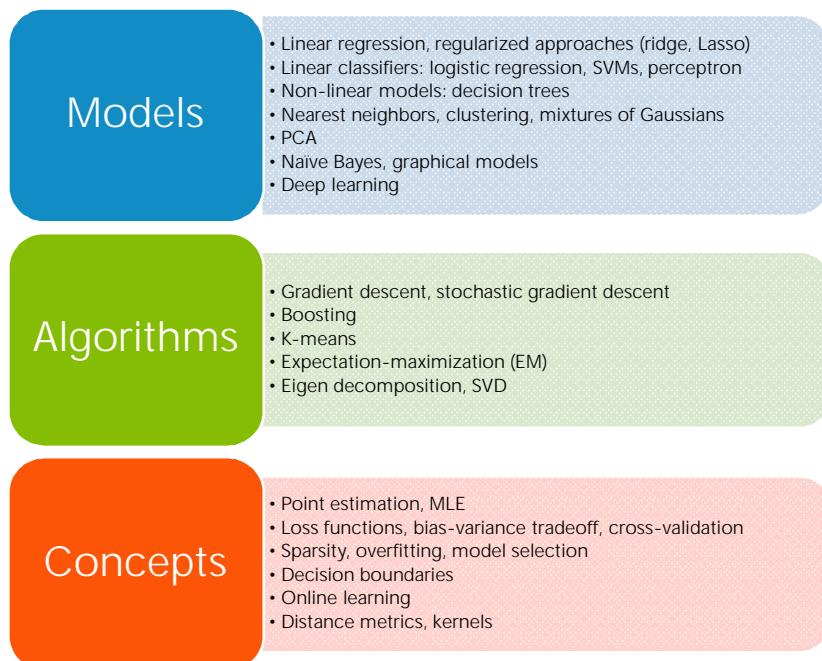


39

©2017 Emily Fox

CSF 446: Machine Learning

Detailed topics



40

©2017 Emily Fox

CSF 446: Machine Learning

Course logistics

Prerequisites

- Formally:
 - Either CSE 326 or CSE 332; either STAT 390, STAT 391, or CSE 312
- Probabilities
 - Distributions, densities, marginalization...
- Basic statistics
 - Moments, typical distributions, regression...
- Algorithms
 - Dynamic programming, basic data structures, complexity...
- Programming
 - Python will be very useful, but we'll help you get started
- We provide some background, but the **class will be fast paced!**
- Ability to deal with "abstract mathematical concepts"

Tutorials

- Python
 - Thurs, Jan 5 (tomorrow!)
 - No 12:30-1:20pm section
- Linear algebra
 - Thurs, Jan 12
- Midterm review
 - Thurs, Feb 2
- Final review
 - Thurs, Mar 9

Communication channels

- Catalyst discussion board
 - For **all non-personal questions**
 - Answering your question will help others
 - Feel free to (and please do!) chime in
- Instructor email list: cse446-staff@cs.washington.edu
 - **Only for personal issues**

Course staff + office hours

Instructor:

- Emily Fox
 - Office hours: [Fridays](#), 10:30-11:30am, Padelford B-305

TAs:

- Dae Hyun Lee
 - Office hours: [Mondays](#), 2-3pm, CSE 021
- Sachin Mehta
 - Office hours: [Wednesdays](#), 2-3pm, CSE 220
- David Wadden
 - Office hours: [Thursdays](#), 2-3pm, CSE 021
- Kaiyu Zheng
 - Office hours: [Tuesdays](#), 2-3pm, CSE 4th floor breakout

Textbooks

- Required textbook:
 - Machine Learning: A Probabilistic Perspective; Kevin Murphy
- Optional Books:
 - Pattern Recognition and Machine Learning; Chris Bishop
 - The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Trevor Hastie, Robert Tibshirani, Jerome Friedman
 - Machine Learning; Tom Mitchell
 - Information Theory, Inference, and Learning Algorithms; David MacKay



Homeworks

Homeworks are **hard**, **start early** 😊

Submission procedure and late policy:

- Use Catalyst to submit homeworks
- **Due before the start of class**
- 33% subtracted per late day
- All homeworks **must be handed in**, even for zero credit

Collaboration policy:

- You may **discuss** the questions
- Each student writes their own answers
- **Write on your homework anyone with whom you collaborate**
- **Each student must write their own code for the programming part**
- **Please don't search for answers on the web, Google, previous years' homeworks, etc.**
 - please ask us if you are not sure if you can use a particular reference

Exams

- Midterm
 - In-class
 - Weds, Feb 8
- Final
 - Finals week
 - Date/time/location TBD
 - (Likely Weds, Mar 15, 8:30-10:20am)

Project

- Projects can be selected from a **list of ideas** or **proposed** based on your interests
 - Make sure you have data available and a clear roadmap...**quarter is short!**
 - Can work as an **individual** or **teams of 2**

Project proposals

- Mon., Feb 6 by 9:30am

Project milestone

- Fri., Feb 24 by 9:30am

Poster session

- Fri., Mar 10, 2-4pm in CSE Atrium

Project report

- Mon., Mar 13 by 9:30am

Grading

Course overload process

Majors:

- Within **60 minutes** of this class ending, fill out the course overload form here: <http://tinyurl.com/hz9sxzd>
- Need code word: **bison**

Non-majors:

- Unfortunately, the course has a huge number of majors waiting and there is almost no chance of getting in
- Please talk to the CSE ugrad advisors in a week or two

Other exciting ML opportunities

- Next year: new non-majors ML course (CSE/STAT 416)
- STAT 435 – Introduction to Statistical Machine Learning
- INFX 574 – Data Science II: Machine Learning and Econometrics
- BIOST 546 – Machine Learning for Biomedical and Public Health Big Data
- DATA 558 – Statistical Machine Learning for Data Scientists

The screenshot shows the Coursera website with a navigation bar at the top. The main content area features a large image of a person working on a laptop with the text "Build Intelligent Applications" and "Master machine learning fundamentals in five hands-on courses.". To the left, a sidebar lists "About This Specialization", "Courses", "Pricing", "Creators", and "FAQs". A blue box highlights the "Machine Learning Specialization" section, which includes a price of "FROM \$39 USD" and a "Enroll Starts Feb 29" button. Below this, a note about financial aid is present. The central content area has a heading "About This Specialization" and a detailed description of the specialization's purpose and structure. It is created by the University of Washington, indicated by their logo and name. Industry partners are also mentioned.

You'll be able to do
amazing things...