Welcome to CSC411

Y. Nikulin and R. Novak, Exploring the Neural Algorithm of Artistic Style

CSC411: Machine Learning and Data Mining, Winter 2018

Michael Guerzhoy and Lisa Zhang

Many slides from Daniel Sheldon

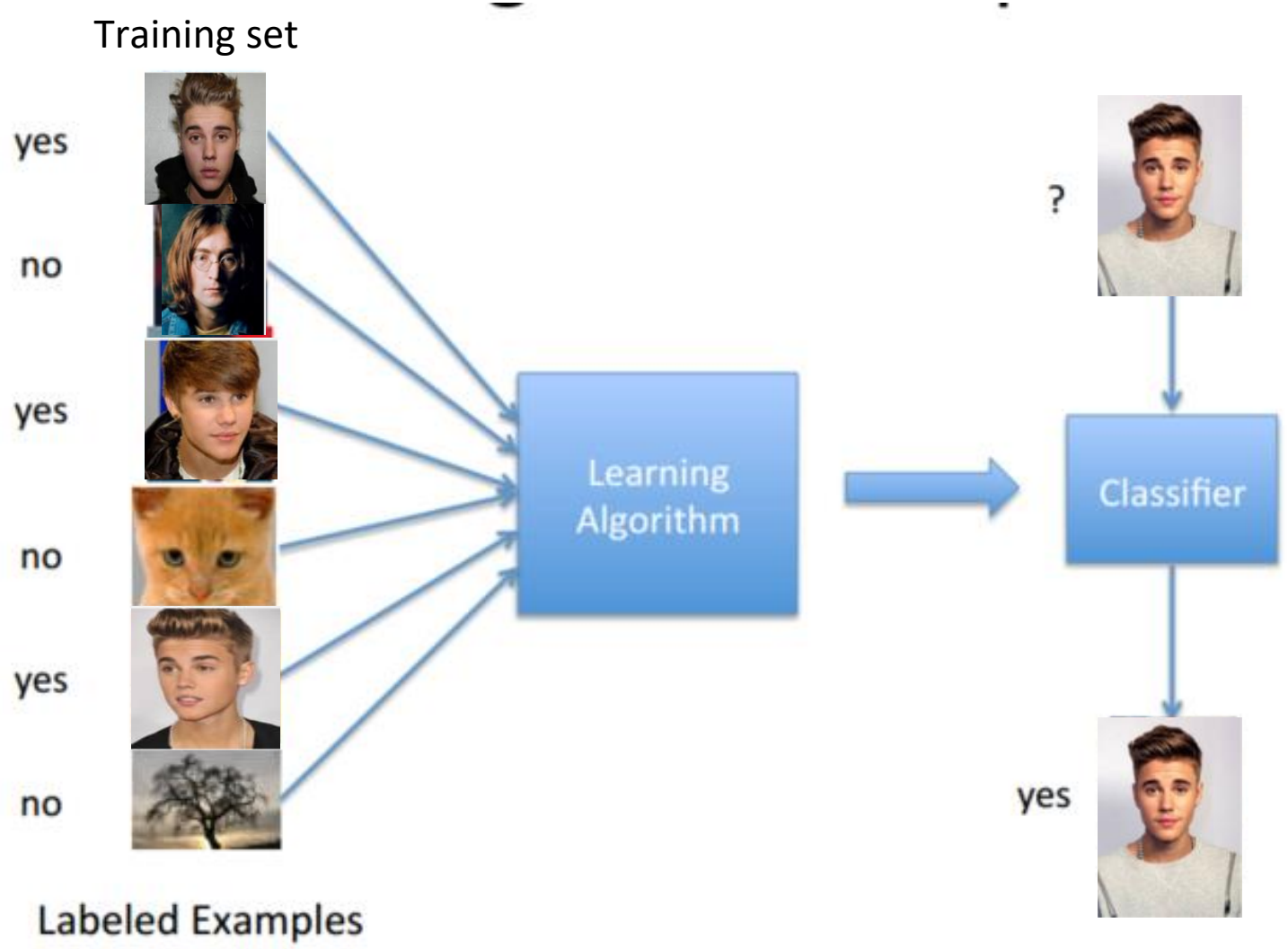
Machine Learning

- Tasks which are hard to solve programmatically:
  - Recognizing faces
  - Recommending movies based on a person's ratings
  - Deciding which web pages are relevant to a given Google search query

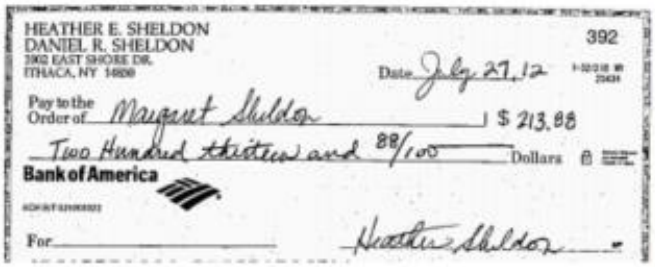
A simple task: Recognizing Justin Bieber

To a computer, this is what Justin Bieber looks like

Machine Learning: Learn From Examples



Machine Learning Applications



Neural Network Applications

ML watches YouTube for three straight days!
(and learns to recognize cats)

http://www.npr.org/2012/06/26/155792609/a-massive-google-network-learns-to-identify

Building High-level Features Using Large Scale Unsupervised Learning

Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Mathieu Devin, Kai Chen, Greg S. Corrado, Jeffrey Dean, and Andrew Y. Ng

Course Goals

- Learn the basic building blocks and the general principles of designing machine learning algorithms
  - Understand the mathematical ideas needed for understanding and designing machine learning/neural networks algorithms
- Learn basic machine learning algorithms
- Learn the methodology of applying machine learning algorithms to data and evaluate their performance
- Apply the principles of machine learning in the context of neural networks and (probably) reinforcement learning

Required math background

- Calculus: derivatives, derivatives as the slope of the function; integrals (a little bit)
- Probability: random variables, expectation, independence
- Linear Algebra: vectors: the dot product, vector norm, vector addition; matrices: matrix multiplication. (Probably: eigenvectors)
- Other topics may be needed, but will be covered in class
- It really helps to *like and be comfortable with* math: you probably won't enjoy this class if you dislike math

Other Logistics

- Course Website:
  - http://www.teach.cs.toronto.edu/~csc411h/winter/
  - All announcements will be made on Piazza
  - Please ask questions on **Piazza** if they are relevant to everyone!

Section	Instructor	Office Hour	Email
Day (T1-3,Th2-3)	Lisa Zhang	Th10-12(BA3219)	lczhang at
Night (Th6-9)	Michael Guerzhoy	M6-7,W6-7(BA3219)	guerzhoy at

Administrative details

- Marking scheme to be finalized and posted within the next few days
- **Tentative** marking scheme
  - 4 projects worth 40%
    - First one to be done individually, others probably can be done in teams of two
  - A midterm worth 15-30%
    - Scheduled for Friday March 2 6pm-8pm (alternate timeslots available in case of documented conflicts)
  - An (undergrad) exam / (grad) project worth about 30-45%
    - Must receive at least 40% on the exam/project (after adjustment) to pass the course

Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

…etc…

Machine Learning: types

Supervised Learning

- Learn to estimate a target y using data x.
  - x : an image
    - y : Justin Bieber?
  - x : text
    - y : Spam?
  - x : house size, location, age, n of bedrooms
    - y : house sell price
- Training set: examples $(x^{(i)}, y^{(i)})$

Machine Learning: types

- Learn something about the datapoints x
  - No targets y
- x: Youtube videos
  - Automatically divide videos into categories ("clustering")
  - *or* Generate images of objects that appear in the videos often
    - Useful for categorizing
    - Useful for automatically learning concepts from videos
- x: User behaviour logs
  - What is the probability that the behaviour would be observed (if it's small, the behaviour might be fraudulent)
- Training set: datapoints $x^{(i)}$

Unsupervised Learning

Machine Learning: types

Reinforcement Learning

- No training set initially
- Get data by interacting with the environment
- Goal: interact with the environment in a way that maximizes the reward you get in the long term
  - Control a robot
    - Try different things, see what gets it to walk the farthest
  - Play Go
    - Play against yourself, try different ways to generate moves, see what strategies allow you to win the most
  - Playing video games

Machine Learning

- What functions are *efficiently* learnable from example data? (ML Theory)
- How can we apply machine learning without compromising data privacy?
- How to make sure machine learning algorithms don't result in unfair discrimination against people?

…etc…

Main focus for most of course:

Supervised Learning

- **Classification**: categorical outputs **y**
- **Regression**: numerical (continuous) outputs **y**

This week

- We will discuss two very different algorithms:
  - K-Nearest Neighbours
  - Linear Regression