

# Introduction

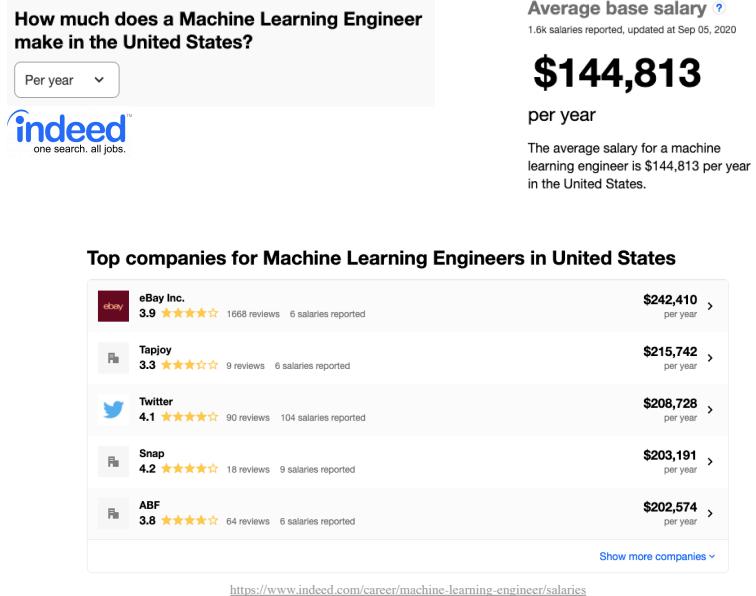
## CSC 461: Machine Learning

Fall 2020

Prof. Marco Alvarez  
University of Rhode Island

## A few quotes

- ▶ “Machine learning is the next Internet”  
✓ (Tony Tether, form. Director, DARPA)
- ▶ “Machine learning is the hot new thing”  
✓ (John Hennessy, President, Stanford)
- ▶ “Machine learning is going to result in a real revolution”  
✓ (Greg Papadopoulos, CTO, Sun)
- ▶ “AI vastly more risky than North Korea”  
✓ (Elon Musk, CEO Tesla, Spaxe X, ...)
- ▶ ...



## Logistics

# Welcome

- Instructor
  - ✓ Prof. Marco Alvarez
- Lectures
  - ✓ MW 3 - 4:15p @ Online
- Office Hours
  - ✓ M 4:30 - 5:30p @ Online
  - ✓ F 4 - 5 @ Online
- Course Website
  - ✓ TBA

# Course goals

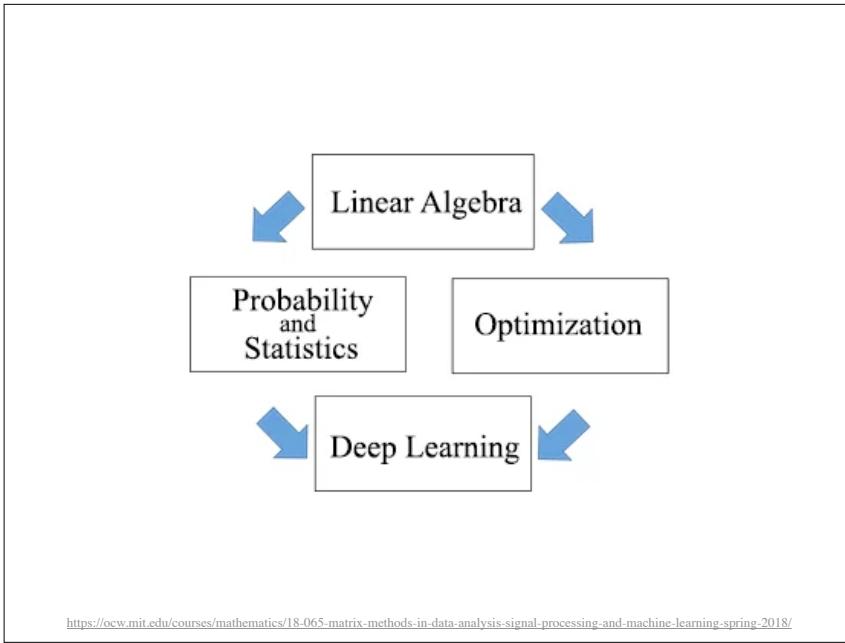
- Understand how ML algorithms work
  - ✓ the learning problem and limitations
  - ✓ theoretical foundations of major techniques
- Be able to develop ML applications
  - ✓ problem design, algorithm/platform choice
- Be able to read ML papers

# Should I take this class?

- Requires more math than traditional CS courses
- Programming experience is required
  - ✓ consider taking this course at a later time if necessary
- Little emphasis on ‘how to use this library’
  - ✓ focus on implementing major algorithms
- High grades require high effort
  - ✓ long and challenging assignments

# Math resources

- Linear Algebra Review and Reference
  - ✓ <http://cs229.stanford.edu/summer2019/cs229-linalg.pdf>
- Review of Probability Theory
  - ✓ <http://cs229.stanford.edu/summer2019/cs229-prob.pdf>
- Computational Linear Algebra for Coders
  - ✓ <https://github.com/fastai/numerical-linear-algebra>
- Mathematics for Machine Learning
  - ✓ <https://gwthomas.github.io/docs/math4ml.pdf>



## Python/Numpy resources

- Google's Python Class
  - ✓ <https://developers.google.com/edu/python/>
- From Python to Numpy
  - ✓ <https://www.labri.fr/perso/nrougier/from-python-to-numpy/>
- Python Numpy Tutorial
  - ✓ <http://cs231n.github.io/python-numpy-tutorial/>

## Tentative topics (order not relevant)

Introduction, Supervised Learning	Decision Trees, Regression Trees
Linear Regression, Polynomial Features	Bagging, Boosting
Linear Classifiers, Loss Functions	Maximum Likelihood Estimation, Naive Bayes
Logistic Regression, Gradient Descent	k-Means, Hierarchical Clustering
Generalization, Bias/Variance, Overfitting/Underfitting, Model Selection	PCA, SVD, Matrix Factorization
Linear SVM, Empirical Risk Minimization, Regularization	Reinforcement Learning
Kernels, Gaussian Processes	Neural Networks
K-NN, Curse of Dimensionality	Convolutional Networks

## Free textbooks



## More textbooks



## Grading

- ▶ Assignments (35%)
  - ✓ ~8 Homework Assignments
- ▶ In-class Presentation (25%)
  - ✓ seminar style
- ▶ Final Project
  - ✓ progress report (10%)
  - ✓ final report (15%)
  - ✓ presentation (15%)

## Final Project

- ▶ Find some new data/problem or find new approach to old data/problem
  - ✓ 2-3 students per group
- ▶ Deliverables
  - ✓ progress report (mid October)
  - ✓ final report (end of semester)
  - ✓ presentation (end of semester)

## Logistics



# Your job

- ▶ Attend lectures (**synchronous**)
- ▶ Participate
  - ✓ in-class and also on Piazza
- ▶ Work hard
  - ✓ read textbooks and papers (schedule is ambitious)
  - ✓ work on your assignments (focus on **excellence** rather than just “getting a good grade”)
  - ✓ this is about developing highly-sought skills and competencies

# The badges game

## Badges data

- ▶ COLT conference 1994
  - ✓ attendees received badges labeled as **positive** or **negative**
- ▶ The author (Haym Hirsh) knew the function that generated the labels
- ▶ Challenge: look at the names and find the hidden function
- ▶ <https://www.seas.upenn.edu/~cis519/fall2019/assets/lectures/lecture-0/game.html>

## Subset of the original dataset

+ Naoki Abe	- Myriam Abramson	+ David W. Aha
- Kamal N. Ali	- Eric Allender	+ Dan Aspinwall
- Chidanand Apte	+ Minoru Asada	+ Lars Ahver
+ Javed Aslam	+ Haralabos Athanassiou	+ Jose L. Balcazar
+ Timothy P. Barber	+ Michael W. Barley	- Cristina Baroglio
+ Peter Bartlett	- Eric Baum	+ Welton Becket
- Shai Ben-David	+ George Berg	+ Neil Berkman
+ Mahesh Bhandaru	+ Bill Bernhard	+ Reinhard Blasig
+ Kevin Blittersdorf	- Annelise Blumer	+ John Boyan
+ Carla Brodsky	+ Nader Bshouty	- Wray Buntine
- Andrej Burago	+ Tom Bylander	+ Bill Byrne
- Claire Cardie	+ Richard A. Caruana	+ John Case
+ Jason Catlett	+ Nicolo Cesa-Bianchi	- Philip Chan
+ Mark Changizi	+ Pang-Chieh Chen	- Zhiqiang Chen
+ Michael Chertkov	- Michael Chien	+ Jeffery Clemons
+ William Cohen	+ David Cohen	- Clark Bain Congdon
- Antoinette Cornejojols	+ Mark W. Craven	+ Robert P. Daley
+ Lindley Darden	- Chris Darken	- Bhaskar Dasgupta
- Brian D. Davidson	+ Michael de la Maza	- Olivier De Vel
- Scott E. Decatur	+ Gerald F. DeJonge	+ Kan Deng
+ Thomas G. Dietterich	+ Michael J. Donahue	+ George A. Drastal
+ Bruce Dumacher	- Christopher Dunn	+ Ian Dunn
- Thomas Elman	+ Tapio Elomaa	+ Susan L. Epstein
+ Bob Evans	+ Claudio Fauchinetti	+ Tom Fawcett
- Usama Fayyad	+ Aaron Feigelson	+ Nicolas Flechter
+ David Finton	+ John Fischer	+ Paul Fischer
+ Seth Flanders	+ Lance Fortnow	- Ameur Foued
+ Judy A. Franklin	+ Younghwan Freeman	+ John Friedmann
+ Michael R. Forst	+ Sean Carroll Gasnacia	+ William Gararch
+ Ricard Gavaldà	+ Melinda T. Gervasio	+ Yolanda Gil
+ David Gillman	- Attilio Giordano	+ Kate Goelz
+ Paul W. Goldberg	+ Sally Goldman	+ Diana Gordon
+ Geoffrey Gordon	+ Jonathan Gratch	+ Leslie Grate
+ William A. Greene	+ Russell Greiner	+ Mark Grobelnik
+ Paul Grunbaum	+ Michael Greenblatt	+ Tom Hadouk
+ Earl S. Harris Jr.	+ David Haussler	+ Matthias Heger
+ Lisa Hellerstein	+ David Helmbold	+ Daniel Hennessy
+ Haym Hirsh	+ Jonathan Hodgson	+ Robert C. Holte
+ Jiarong Hong	- Chun-Nan Hsu	+ Kazushi Ikeda
+ Masayuki Inaba	- Drago Indic	+ Nitin Indurkha
+ Jeff Jackson	+ Sameer Jain	+ Wolfgang Jankó
+ Klemens Janke	+ Michael I. Jordan	+ George Jannink
+ Randolph Jones	- Michael J. Kearns	+ Leslie Pack Kaelbling
+ Bala Kalyanasundaram	- Thomas E. Kammerer	- Grigoris Karakoulas
+ Michael Kearns	+ Neela Khan	+ Roni Kharon
+ Dennis F. Kibler	+ Jorg-Uwe Kietz	- Efin Kinber
- Jyrki Kivinen	- Emanuel Knill	- Craig Knoblock
+ Ron Koenig	+ Pauline Koiran	+ Moshe Koppel
+ Daniel Kortenkamp	+ Matevz Kovacic	+ Stephan Kramer
+ Martinch Krikis	- Martin Kümerer	+ Eyal Kushilevitz
- Stephen Ewuk	+ Wai Lam	+ Ken Lang
- Steffen Lange	+ Pai Langley	+ Mary Soon Lee
+ Wee Sun Lee	+ Moshe Leshno	+ Long-Ji Lin

## Lets play ...

- ▶ Analyze the training data ...
- ▶ Lets calculate your accuracy on these names ...

Brian Tester  
Lyle H. Ungar  
Paul Vitanyi  
Gary Weiss  
Bradley L. Whitehall  
Janusz Wnek  
Holly Yanco  
Jean-Daniel Zucker

Chen K. Tham  
Paul Utgoff  
Xuemei Wang  
Sholom Weiss  
Alma Whitten  
Kenji Yamanishi  
John M. Zelle  
Darko Zupanic

## Can we use a computer?

- ▶ We could extract features (attributes/characteristics) then use ML
  - ✓ length
  - ✓ number of vowels
  - ✓ number of consonants
  - ✓ number of dots
  - ✓ number of words
  - ✓ number of whitespaces
  - ✓ vowel/consonant ratio
  - ✓ length even/odd?
  - ✓ starts with a vowel?
  - ✓ is second letter a vowel?
  - ✓ ...