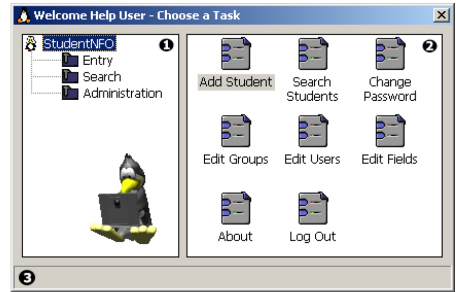
Machine Learning

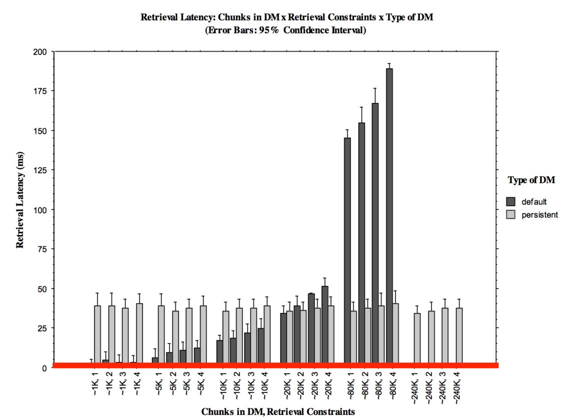
Nate Derbinsky
Assistant Professor
Computer Science and Networking

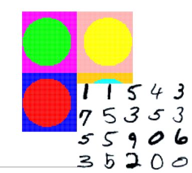
About Me

- Founded a computer consulting business in high school

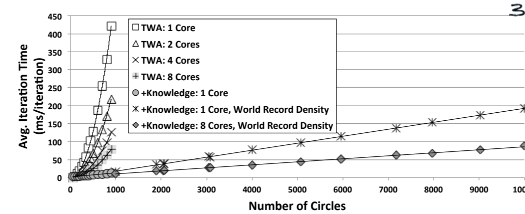


- PhD from University of Michigan (**Go Blue!**)





- Imagineer with Disney Research, Boston



What is Machine Learning (ML)?

The study/construction of algorithms that can learn from data

The study of algorithms that improve their performance **P** at some task **T** with experience **E**
– Tom Mitchell (CMU)

Fusion of algorithms, artificial intelligence, statistics, optimization theory, visualization, …

Natural Language Processing (NLP)

Modern NLP algorithms are typically based on statistical ML

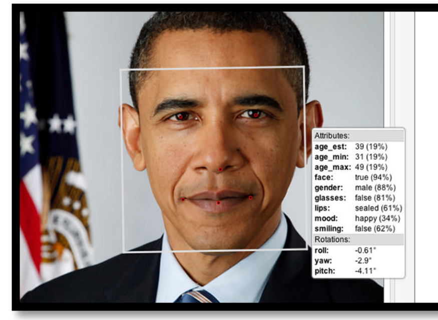
<u>Applications</u>

- Summarization
- Machine Translation
- Speech Processing
- Sentiment Analysis

…

Computer Vision

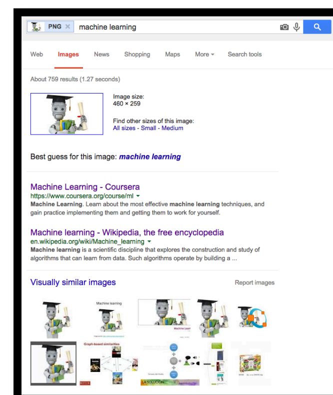
Methods for acquiring, processing, analyzing, and understanding images



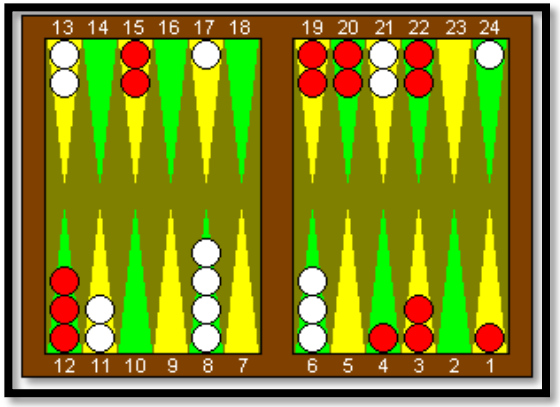
<u>Applications</u>

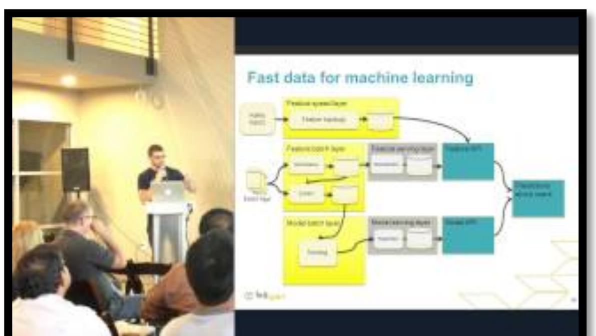
- Image search
- Facial recognition
- Object tracking
- Image restoration

…



Games, Robotics, Medicine, Ads, …





Different Motivations

Position	Salary*
Data Scientist	**$118,709**
Machine Learning Engineer	**$112,500**
Software Engineer	$90,374

"A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician."
– Josh Blumenstock (UW)

"Data Scientist = statistician + programmer + coach + storyteller + artist"
– Shlomo Aragmon (Ill. Inst. of Tech)

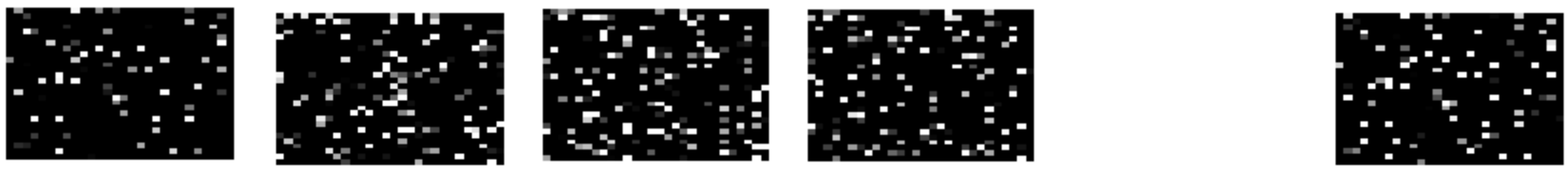
*glassdoor.com, National Avg

Types of Problems

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Training *Testing*



α β β γ ... ?

...

MNIST

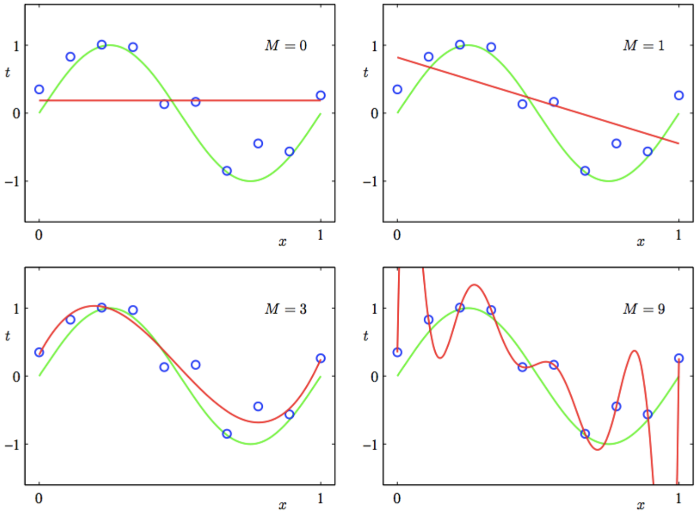
Supervised Tasks

Classification

Binary vs. multi-class



Regression



<u>Issues</u>

- Feature selection
- Overfitting
  – Regularization, cross validation

Common Algorithms

- Nearest Neighbor (kNN)
- Decision Tree learning (e.g. ID3, C4.5, RF)
- Support Vector Machine (SVM)
- Neural Networks
  - Backpropagation
  - Deep learning

kNN

Training

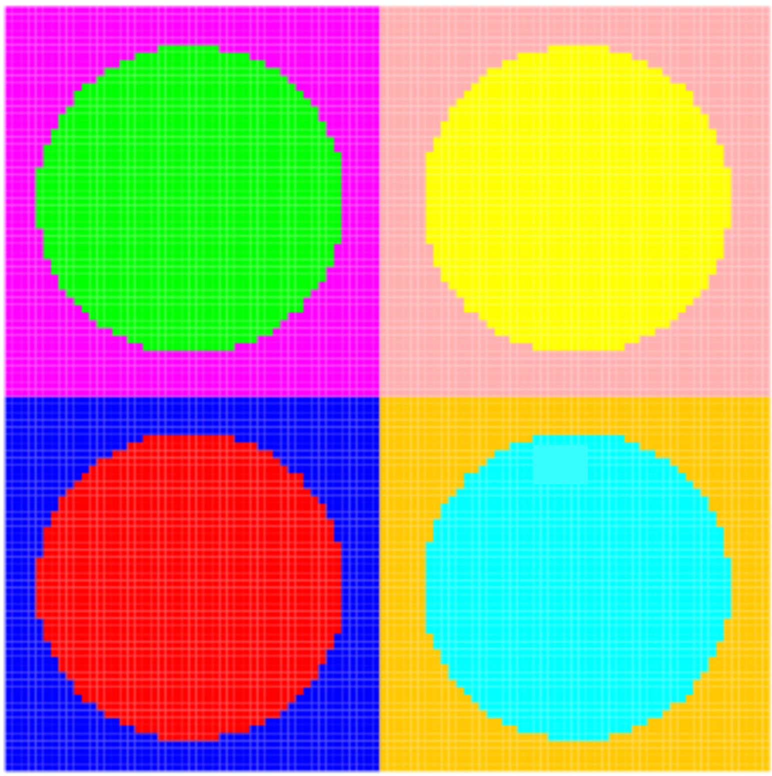
- Store all examples

Testing

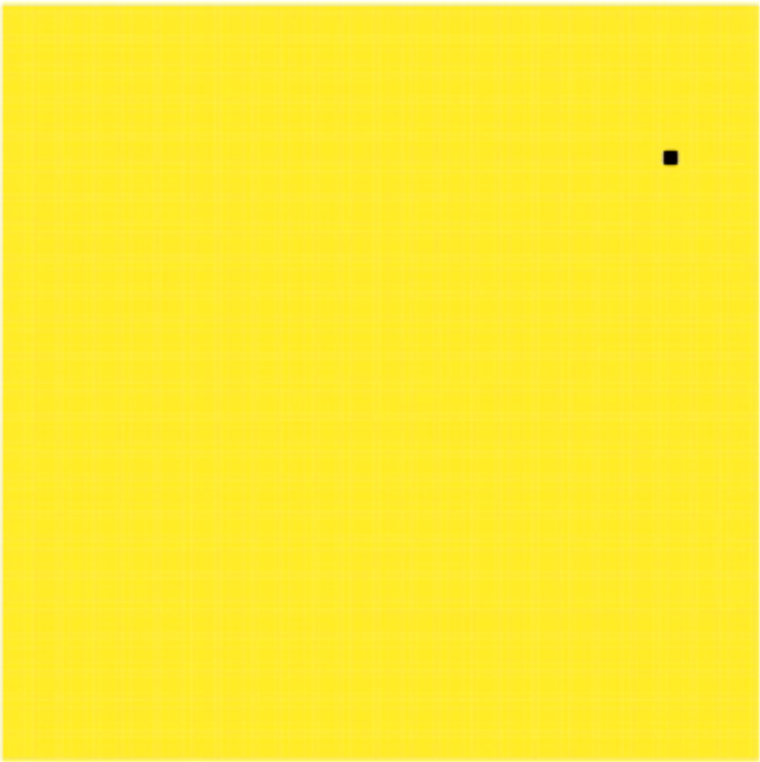
- Find the nearest *k* neighbors to target
  - Via distance function
- Vote on class

2D Multiclass Classification

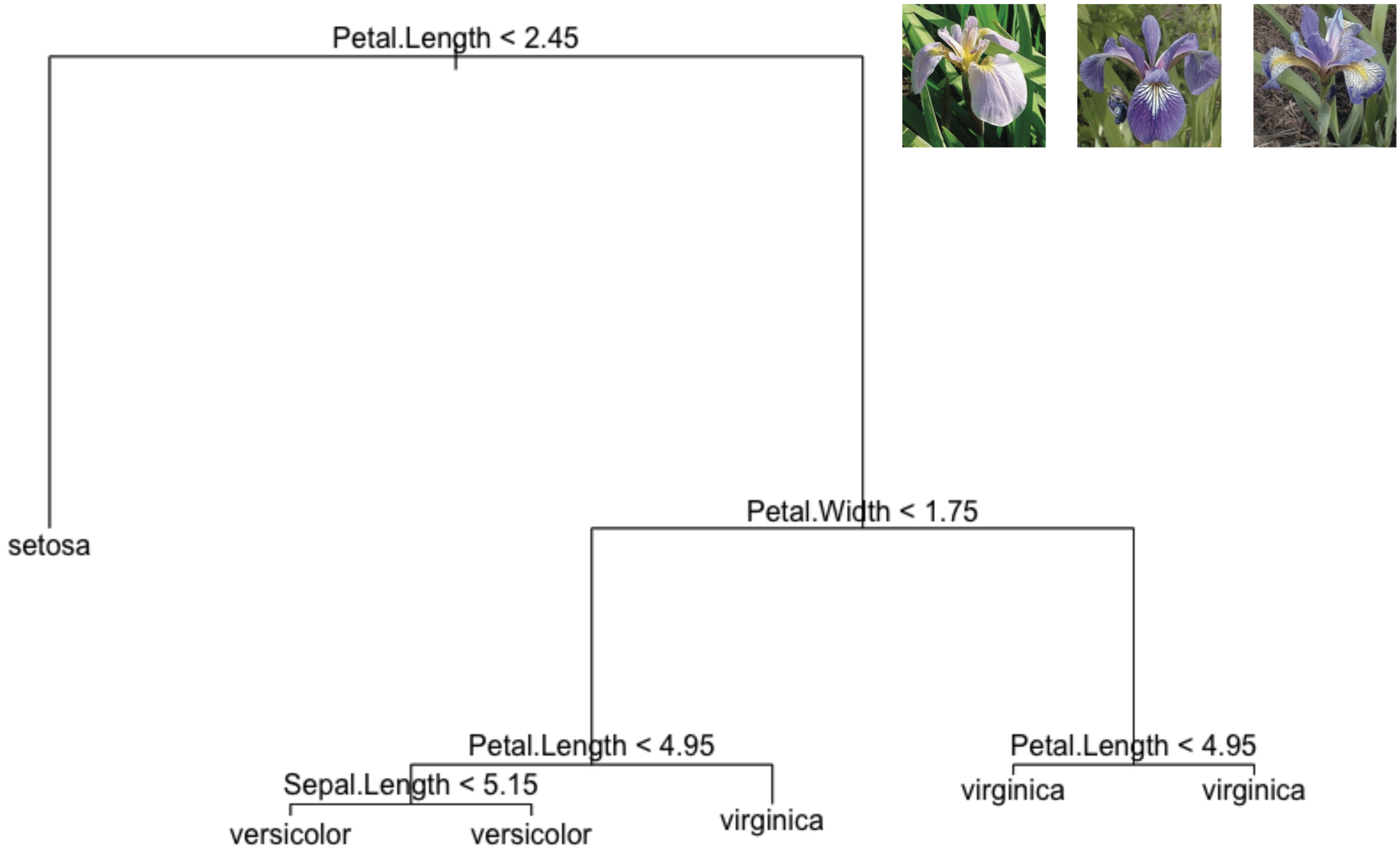
Boundary Tree



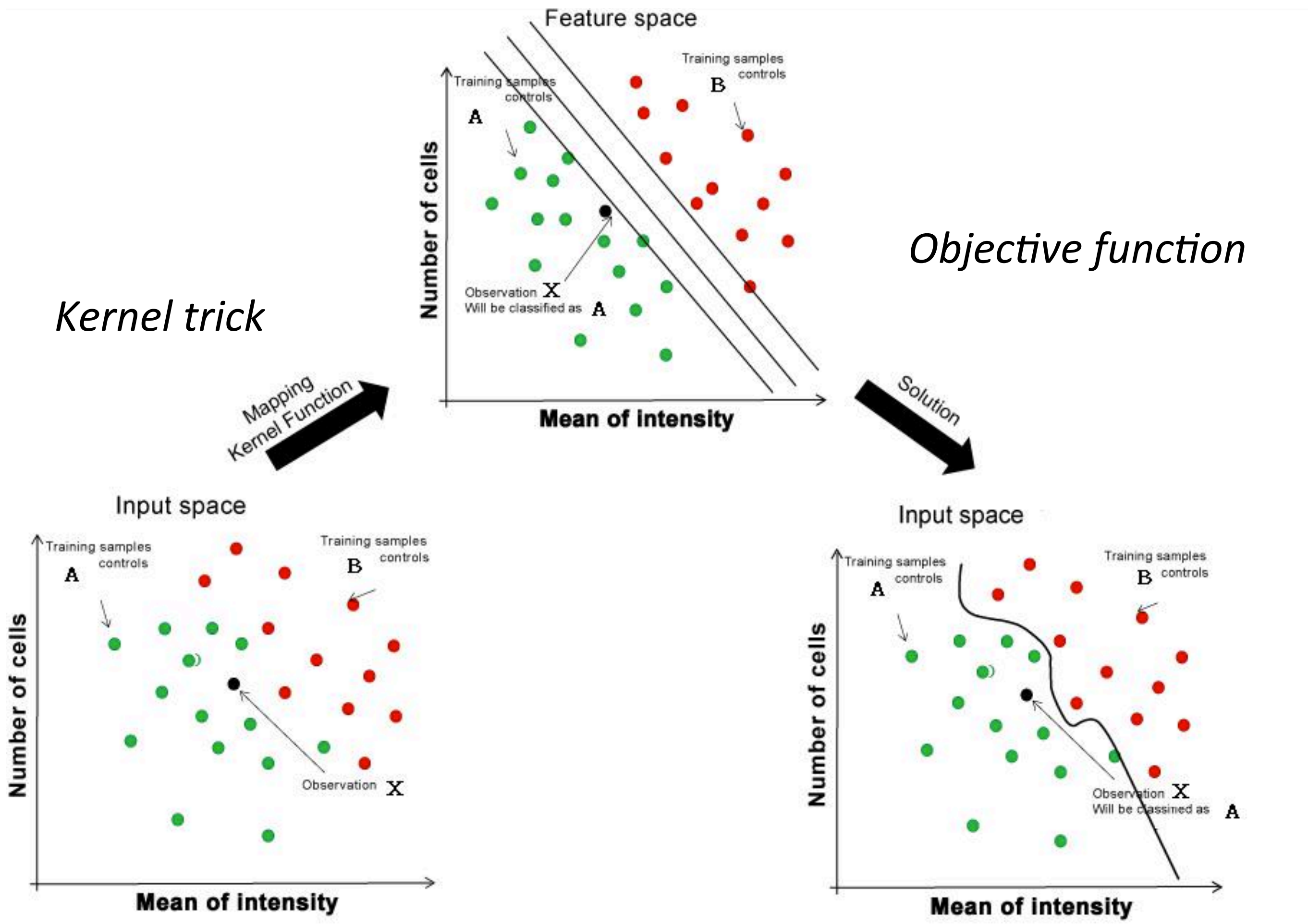
1-NN via Linear Scan



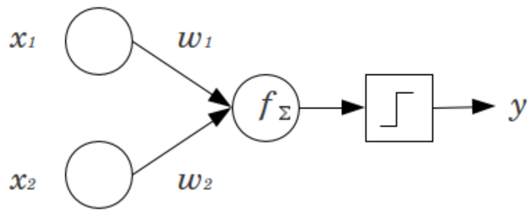
Decision Trees/Forests



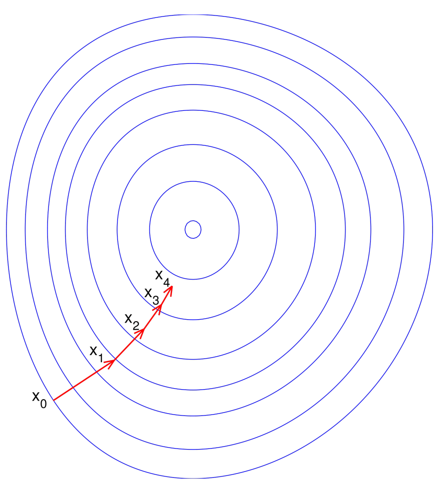
Support Vector Machine (SVM)



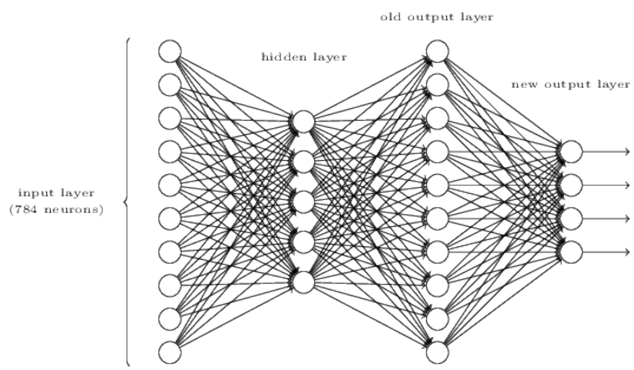
Artificial Neural Networks (ANN)



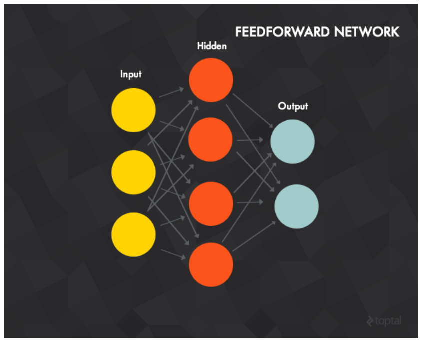
Perceptron
Linear classifier



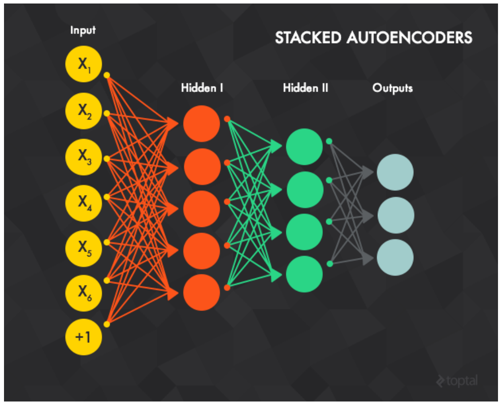
Gradient descent



Backpropagation



Feedforward vs.
Recurrant



Deep Architectures
Vanishing Gradient

Unsupervised Learning

No right answer, find "interesting structure"

Tasks

- Clustering
- Dimensionality reduction
- Density estimation
- Discovering graph structure
- Matrix completion

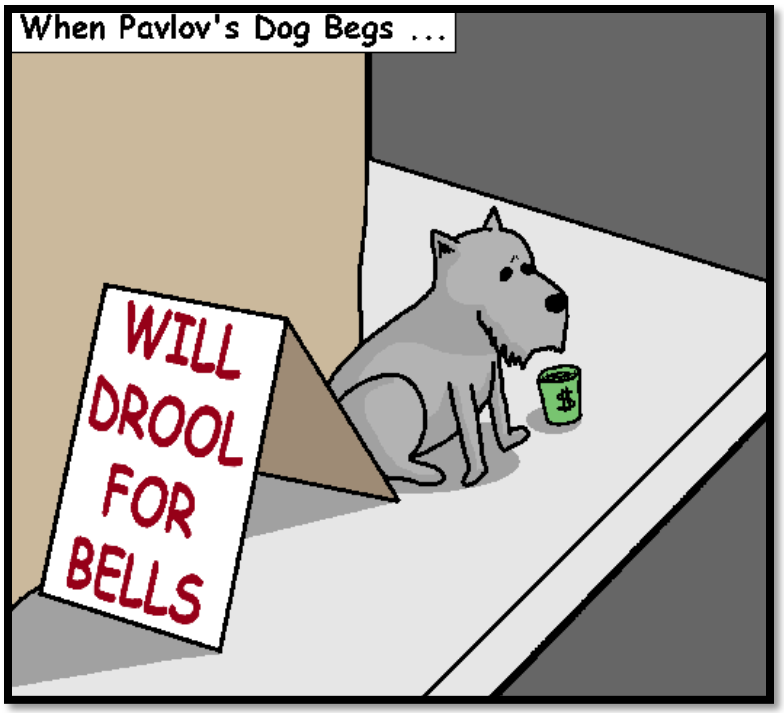
Common Algorithms

- k-Means Clustering
  http://web.stanford.edu/~kvmohan/kmeans-voronoi/kmeans.html

- Collaborative Filtering
- Principle Component Analysis (PCA)
- Expectation Maximization (EM)
- Neural Networks (e.g. RBM)

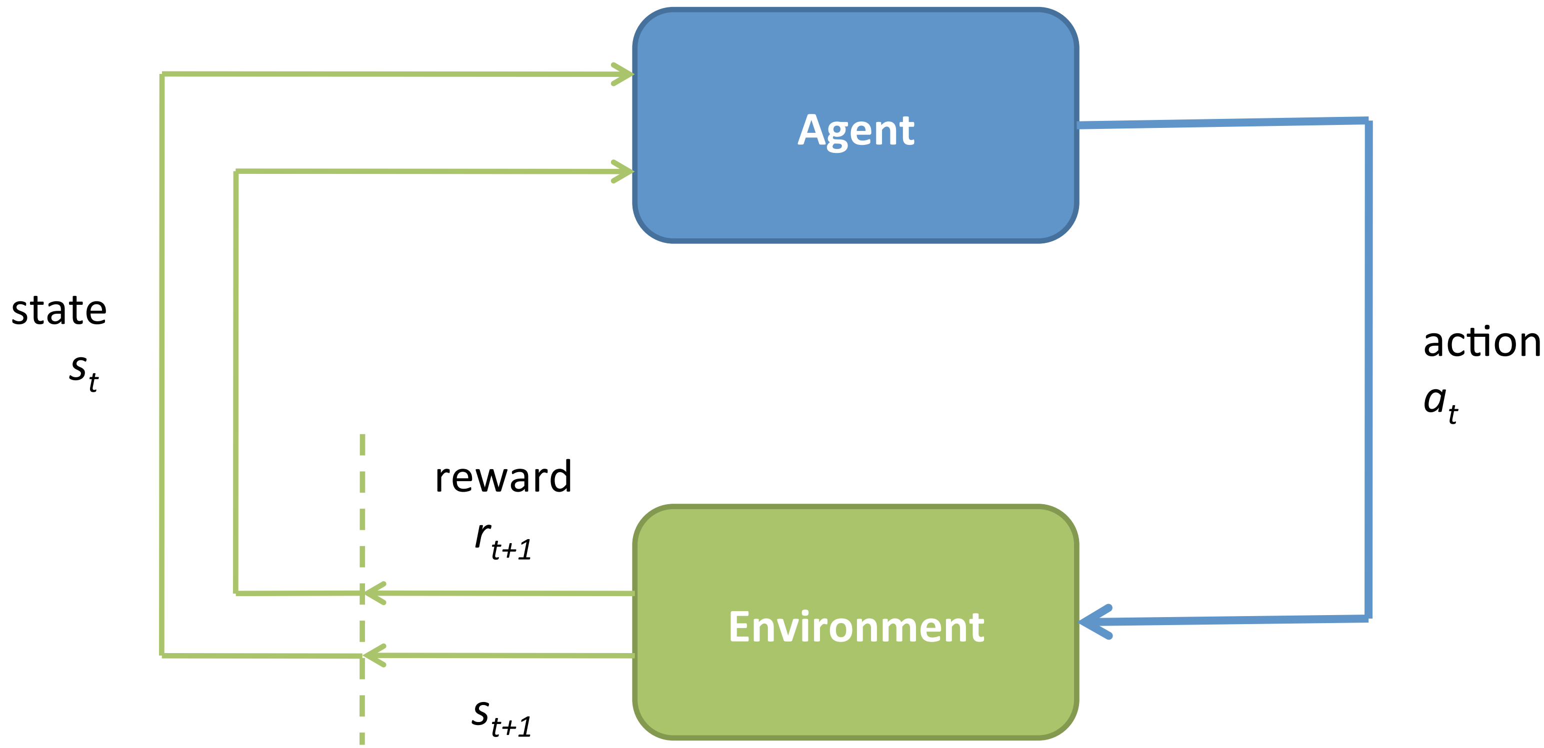
Reinforcement Learning (RL)

Choose actions to maximize future reward



The RL Cycle

Issues. credit assignment, exploration vs. exploitation, reward function, …



Temporal Difference (TD) Learning

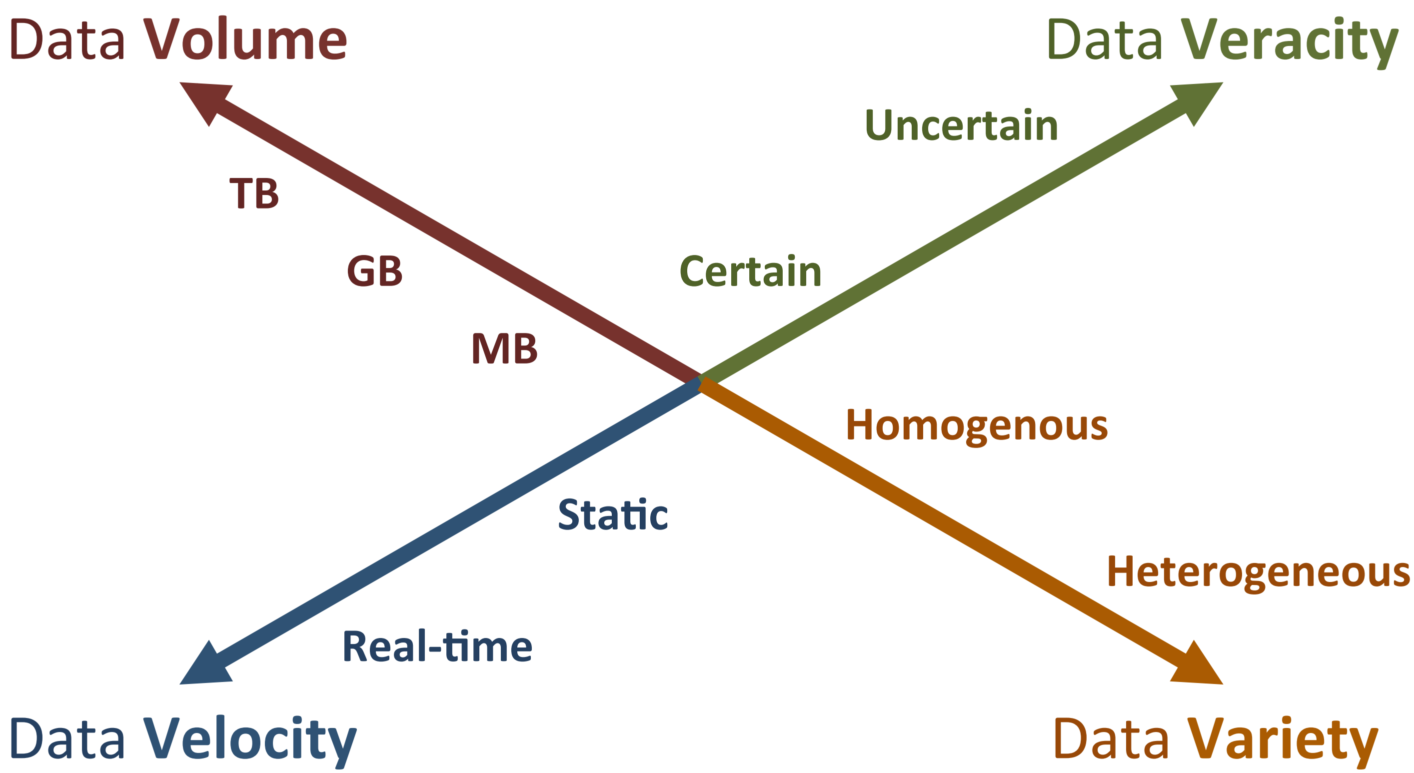
$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

- Evidence that some neurons (dopamine) operate similarly
- Lead to world-class play via TD-Gammon (neural network trained via TD-learning)

Challenges

- Big Data
  - Parametric vs. Nonparametric
- Curse of Dimensionality
- No Free Lunch

Big Data – The Four V's



Parametric algorithm: model does not grow with data size

The Curse of Dimensionality

“Various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience.” – Wikipedia

- Memory requirement increases
- Required sampling increases
- Distance functions become less useful

…

No Free Lunch

- There is no universally best model – a set of assumptions that works well in one domain may work poorly in another

- We need many different models, and algorithms that have different speed-accuracy-complexity tradeoffs

Some Nice Starting Points

Books
- “Machine Learning in Action”
- “Data Mining: Practical Machine Learning Tools and Techniques”
- “Pattern Recognition and Machine Learning”
- “Machine Learning: A Probabilistic Perspective”

Courses
- Fall 2015 @ WIT (MATH+COMP)!
- Coursera: https://www.coursera.org/course/ml
- CMU: http://www.cs.cmu.edu/~tom/10701_sp11/
- UW: https://class.coursera.org/datasci-002/lecture

Sites
- DataTau/Kaggle
- MNIST/UCI Machine Learning Repository
- DataQuest.io
- deeplearning.net
- neuralnetworksanddeeplearning.com

Academia
- E/ICML
- NIPS
- AAAI
- UAI

Thank You :)

Questions?

Nate Derbinsky
Dobbs 140
derbinskyn@wit.edu
http://derbinsky.info