Machine Learning 2

DS 4420 / ML 2 / Spring 2020

Hellos, logistics, etc

Byron C Wallace

Some slides derived from Jan-Willem van Meent

Course website:

https://course.ccs.neu.edu/ds4420sp20/

What will be covered?

- A bit tricky actually; you’re all coming in with different backgrounds (and different variants of ML1)

What will be covered?

- A bit tricky actually; you're all coming in with different backgrounds (and different variants of ML1)
- Doing my best to minimize overlap in content, but some is inevitable

What will be covered?

- A bit tricky actually; you're all coming in with different backgrounds (and different variants of ML1)
- Doing my best to minimize overlap in content, but some is inevitable
- In short: A focus on un-/semi-supervised learning and advanced techniques

First half of class (through mid-term)

https://course.ccs.neu.edu/ds4420sp20/

(partial?) review for some of you

Meeting	Topic(s)	readings
1/7 (t)	*Logistics, overview*	
1/9 (r)	*Math Review*	Math for ML, Part 1: 5-5.5, 6-6.5
1/14 (t)	*MLE, MAP, and graphical models*	Math for ML, Part 2: 8.3, 8.4, 8.5
1/16 (r)	*Neural networks / backprop*	A Course in Machine Learning, Ch. 10
1/21 (t)	*Clustering I*	Elements of Statistical Learning, 14--14.6; (optional) CIML 11.3
1/23 (r)	*Clustering II → Mixture models and EM*	Elements of Statistical Learning, 14.6--14.9; MML, Part 2: 11
1/28 (r)	*Topic modeling I*	Applications of Topic Models (Boyd-Graber, Hu, Mimno)
1/30 (r)	*Topic modeling II*	Applications of Topic Models (Boyd-Graber, Hu, Mimno)
2/4 (t)	*Dimensionality reduction I*	Math for ML, Part 2: 10
2/6 (r)	*Dimensionality reduction II / Auto-encoders*	t-SNE paper
2/11 (t)	*"Self-supervision"; Learning to embed*	
2/13 (r)	*Structured prediction I*	A Course in Machine Learning, Ch 17
2/18 (t)	*Structured prediction II*	A Course in Machine Learning, Ch 17
2/20 (r)	*No class*	
2/25 (t)	*Review*	
2/27 (r)	*Midterm exam*	

Second half…

https://course.ccs.neu.edu/ds4420sp20/

3/10 (t)	*Transformers*		
3/13 (r)	*Fairness and bias*	A Course in Machine Learning, Ch. 8	Project proposal due
3/17 (t)	*Project pitches and feedback*		In class project pitches!
3/19 (r)	*Interpretability*		
3/24 (t)	*Active learning*		HW4 DUE
3/26 (r)	*"Green" AI*		
3/31 (t)	*Reinforcement learning I*		
4/2 (r)	*Reinforcement learning II*		
4/7 (t)	Final project presentations I		Presentations!
4/9 (r)	Final project presentations II		Presentations!
4/14 (t)	No class (final write-ups due)		FINAL PROJECT WRITE-UPS DUE!

3 Aspects of ML

Data Types

- Sets
- Matrices / Tables
- Graphs
- Time series
- Sequences
- Text
- Images

3 Aspects of ML

Data Types

- Sets
- Matrices / Tables
- Graphs
- Time series
- Sequences
- Text
- Images

Methods

- Association Rules
- Dimensionality Reduction
- Regression
- Classification
- Clustering
- Topic Models
- Bandits

3 Aspects of ML

Data Types

- Sets
- Matrices / Tables
- Graphs
- Time series
- Sequences
- Text
- Images

Methods

- Association Rules
- Dimensionality Reduction
- Regression
- Classification
- Clustering
- Topic Models
- Bandits

Tasks

- Exploratory Analysis
- Game playing
- Prediction

Machine Learning Methods

Supervised Learning

Given *labeled* examples, learn to make predictions for *unlabeled* examples. *Example:* Image classification.

Machine Learning Methods

Supervised Learning

Given *labeled* examples, learn to make predictions for *unlabeled* examples. *Example:* Image classification.

Unsupervised Learning

Given *unlabeled* examples learn to identify structure. *Example:* Community detection in social networks.

Machine Learning Methods

Supervised Learning

Given *labeled* examples, learn to make predictions for *unlabeled* examples. *Example:* Image classification.

Unsupervised Learning

Given *unlabeled* examples learn to identify structure. *Example:* Community detection in social networks.

Reinforcement Learning

Learn to take *actions* that maximize future *reward. Example:* Targeting advertisements.

Regression

Goal: Predict a *Continuous* Label

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93	16.5
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10	18.9
10	0.22489	12.5	7.87	0.0	0.524	6.377	94.3	6.3467	5.0	311.0	15.2	392.52	20.45	15.0

Boston Housing Data (source: UCI ML datasets)

https://archive.ics.uci.edu/ml/datasets/Housing

Regression

Target Variable

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93	16.5
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10	18.9
10	0.22489	12.5	7.87	0.0	0.524	6.377	94.3	6.3467	5.0	311.0	15.2	392.52	20.45	15.0

MEDV: Median value of owner-occupied homes in $1000's

Regression

Features

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93	16.5
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10	18.9
10	0.22489	12.5	7.87	0.0	0.524	6.377	94.3	6.3467	5.0	311.0	15.2	392.52	20.45	15.0

Real-valued

CRIM: per capita crime rate by town

Regression

Features

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93	16.5
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10	18.9
10	0.22489	12.5	7.87	0.0	0.524	6.377	94.3	6.3467	5.0	311.0	15.2	392.52	20.45	15.0

Discrete / Categorical

CHAS: Charles River variable
(= 1 if tract bounds river; 0 otherwise)

Regression

Features

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93	16.5
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10	18.9
10	0.22489	12.5	7.87	0.0	0.524	6.377	94.3	6.3467	5.0	311.0	15.2	392.52	20.45	15.0

Hand-Engineered

DIS: weighted distances to five Boston employment centers

Hmmm…

CRIM: Per capita crime rate by town
ZN: Proportion of residential land zoned for lots over 25,000 sq. ft
INDUS: Proportion of non-retail business acres per town
CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
NOX: Nitric oxide concentration (parts per 10 million)
RM: Average number of rooms per dwelling
AGE: Proportion of owner-occupied units built prior to 1940
DIS: Weighted distances to five Boston employment centers
RAD: Index of accessibility to radial highways
TAX: Full-value property tax rate per $10,000
PTRATIO: Pupil-teacher ratio by town
B: $1000(Bk - 0.63)^2$, where Bk is the proportion of [people of African American descent] by town
LSTAT: Percentage of lower status of the population
MEDV: Median value of owner-occupied homes in $1000s

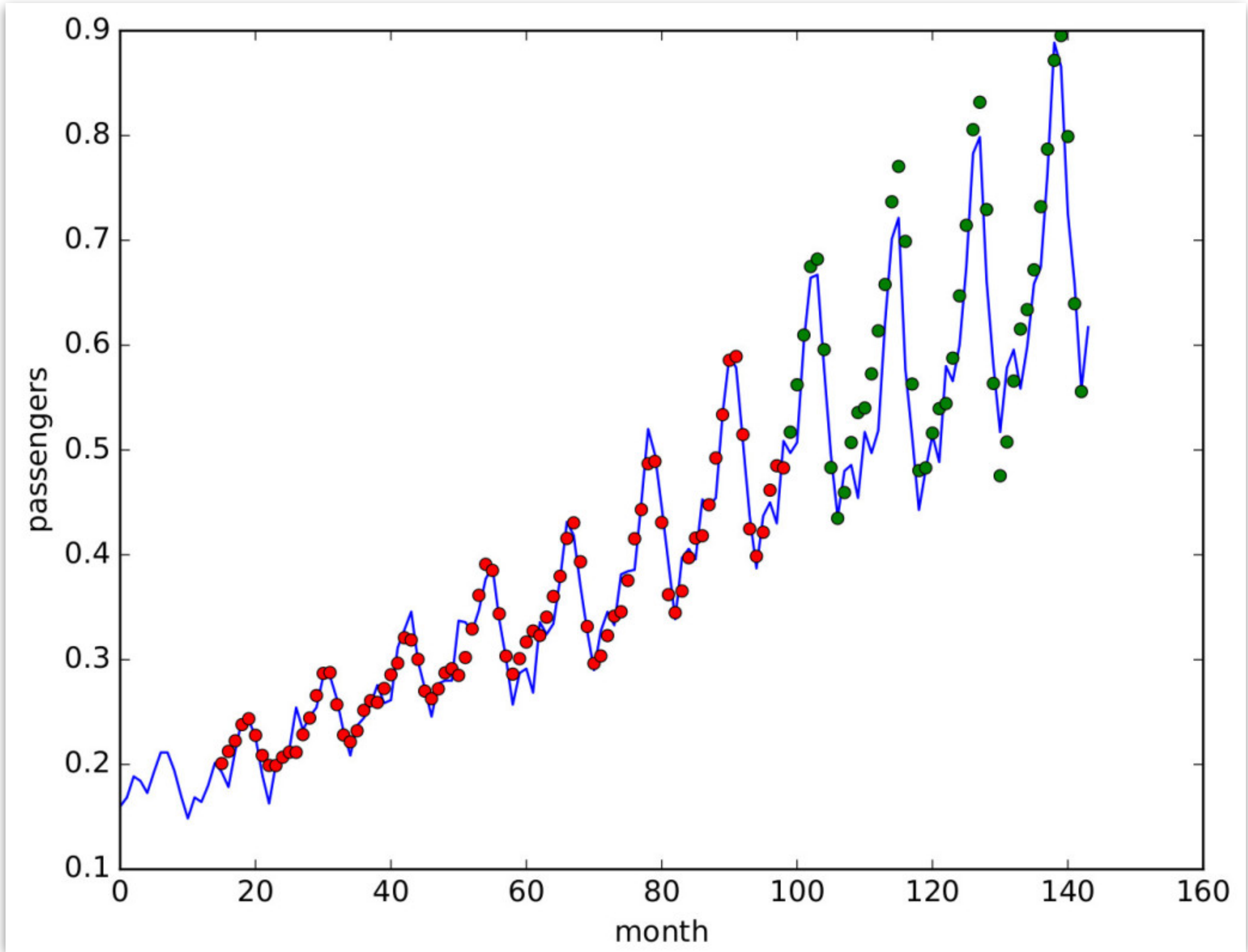
Hmmm…

CRIM: Per capita crime rate by town
ZN: Proportion of residential land zoned for lots over 25,000 sq. ft
INDUS: Proportion of non-retail business acres per town
CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
NOX: Nitric oxide concentration (parts per 10 million)
RM: Average number of rooms per dwelling
AGE: Proportion of owner-occupied units built prior to 1940
DIS: Weighted distances to five Boston employment centers
RAD: Index of accessibility to radial highways
TAX: Full-value property tax rate per $10,000
PTRATIO: Pupil-teacher ratio by town
B: $1000(Bk - 0.63)^2$, where Bk is the proportion of [people of African American descent] by town
LSTAT: Percentage of lower status of the population
MEDV: Median value of owner-occupied homes in $1000s

Regression

Goal: Use past labels (red) to learn trends that generalize to future data points (green)

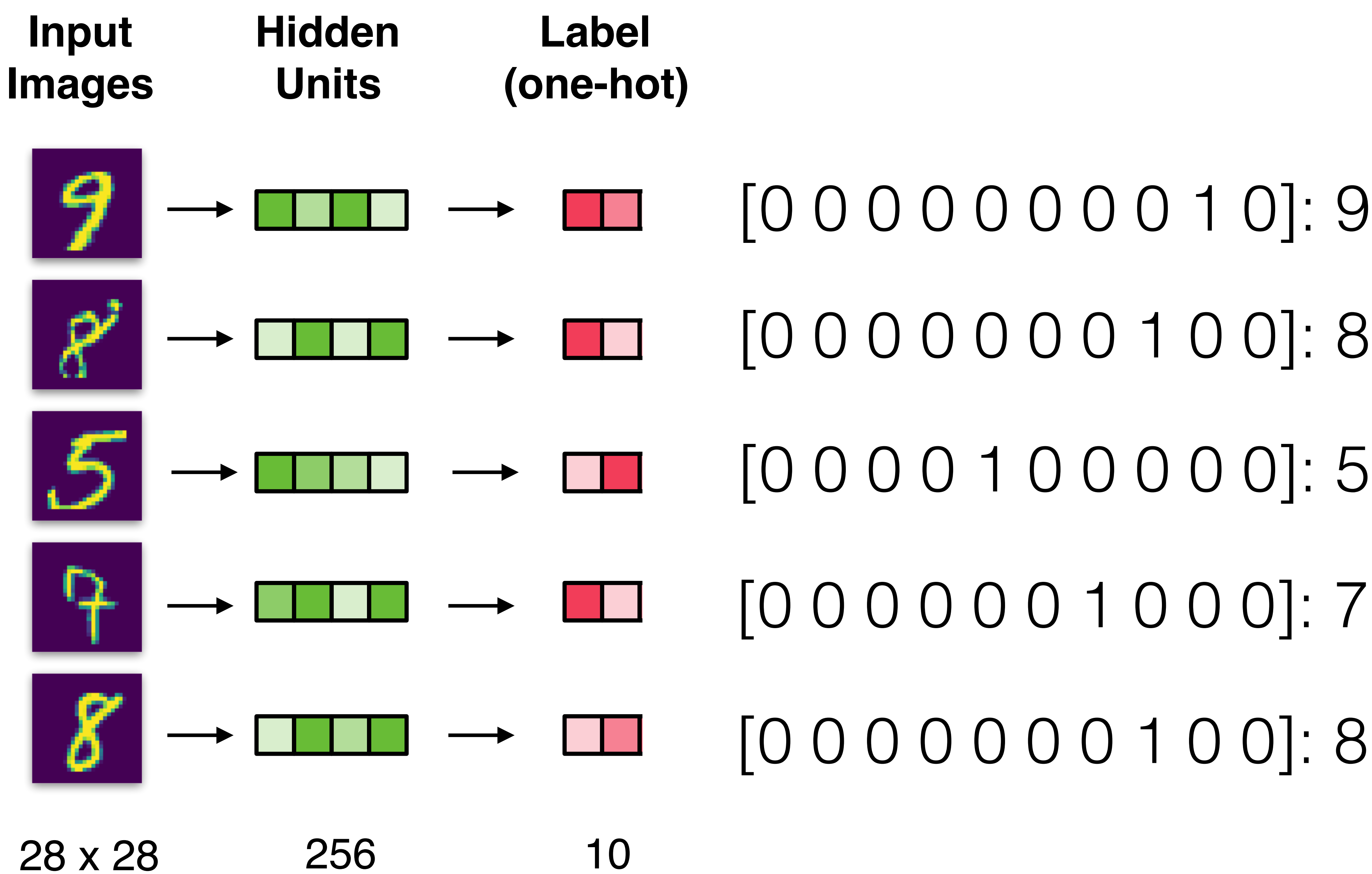
Time-series Data



source: https://am241.wordpress.com/tag/time-series/

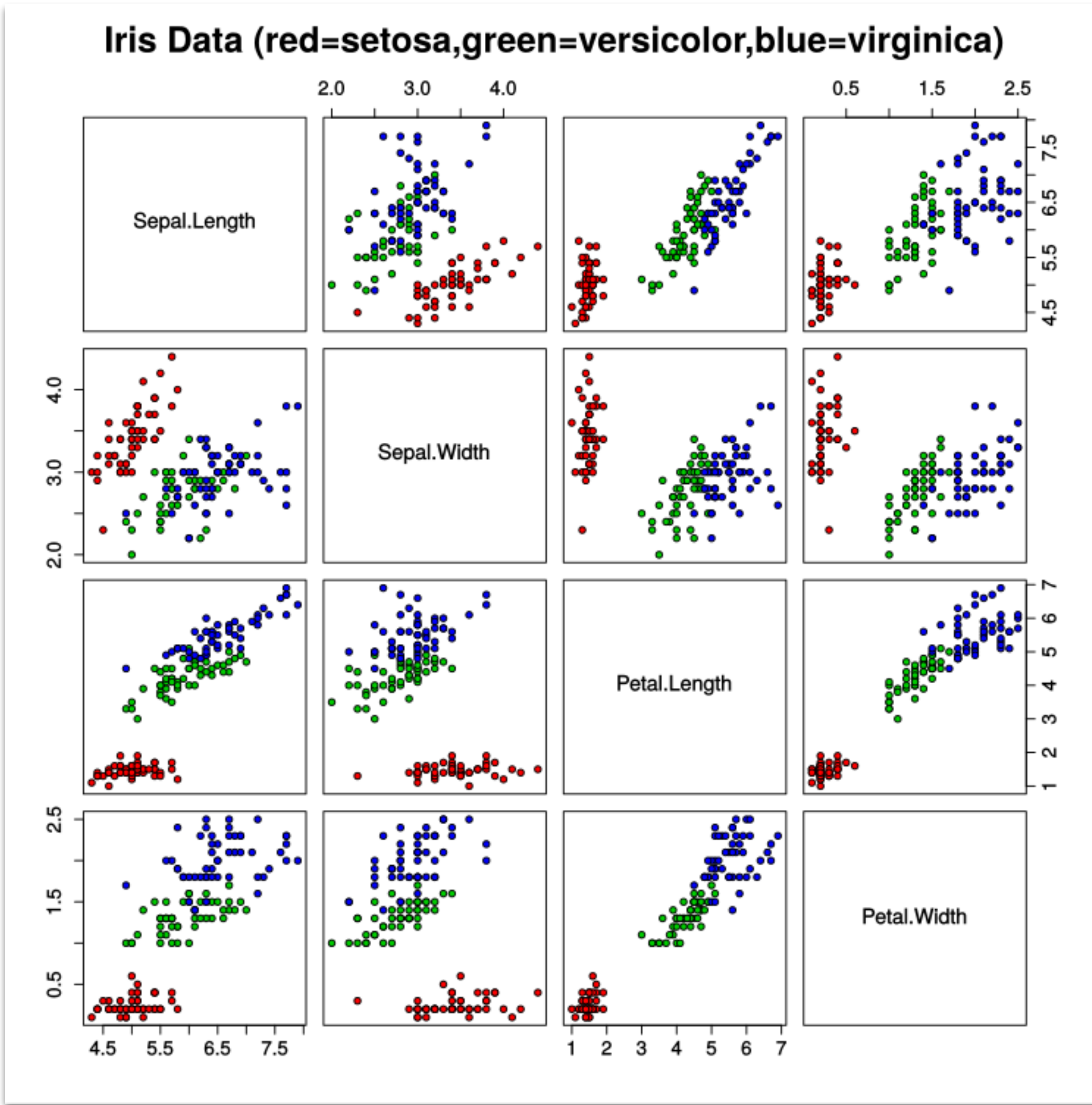
Classification

Goal: Predict a *discrete* label.



Classification

Example: Iris Data



Iris Setosa

Iris versicolor

Iris virginica

https://en.wikipedia.org/wiki/Iris_flower_data_set

Let’s run through a quick practical refresher on classification.

Navigate to blackboard to *in class exercise 0* and grab *refresher.ipynb.*

Unsupervised Learning

Goal: Can we make predictions in absence of labels?

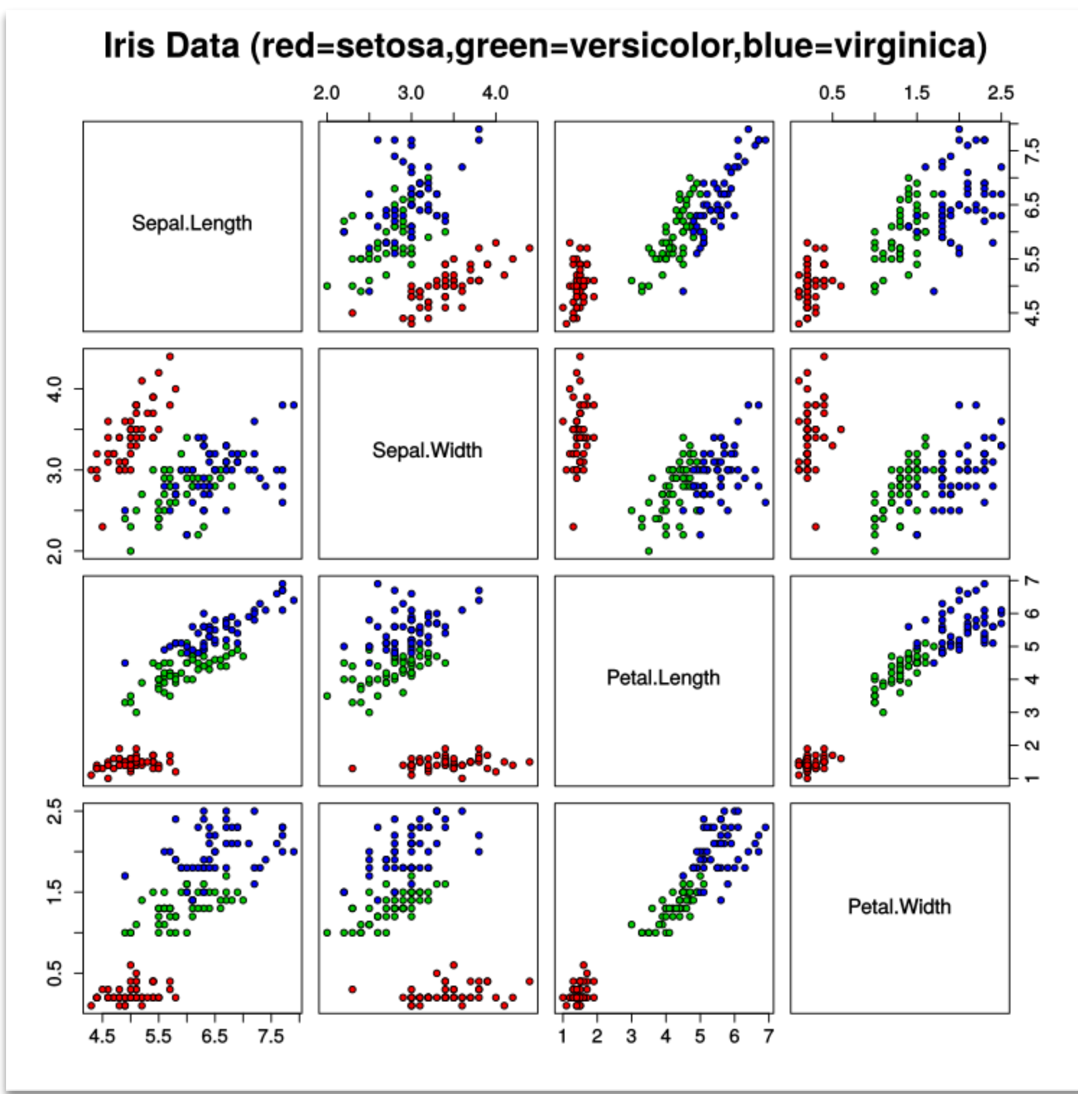
Prominent examples:

- Dimensionality Reduction
- Clustering
- Topic Modeling

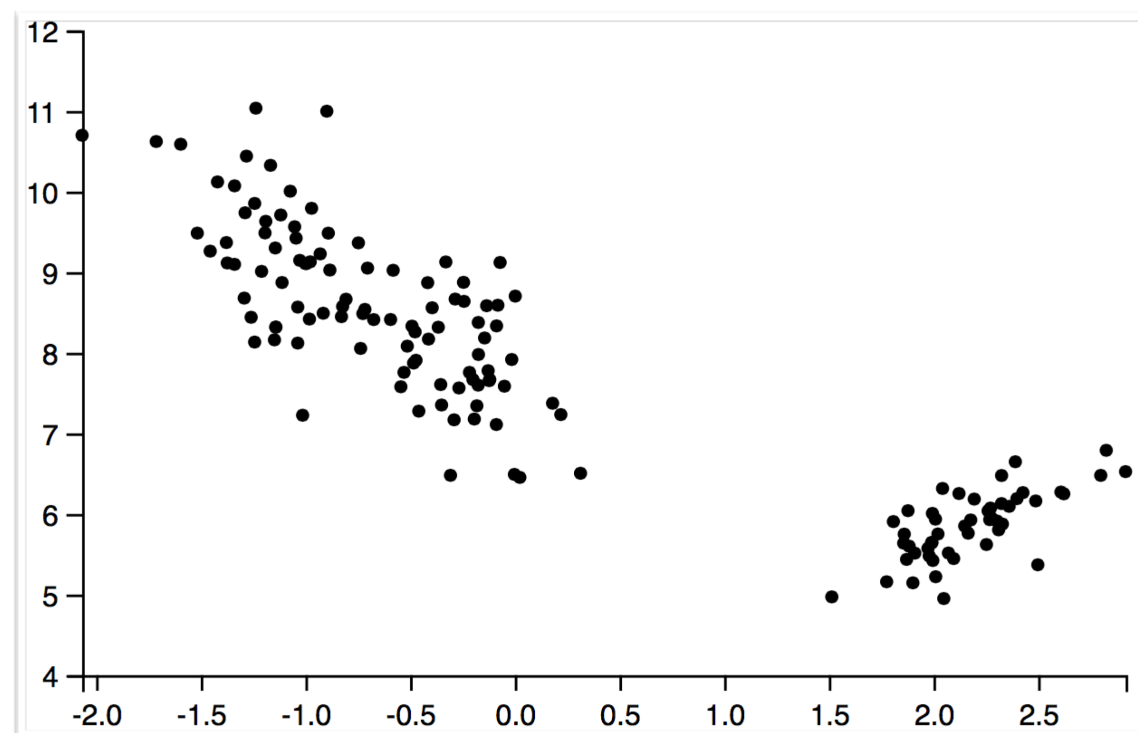
Dimensionality Reduction

Goal: Map high dimensional data onto lower-dimensional data in a manner that preserves *distances/similarities*

Original Data (4 dims)



Projection with PCA (2 dims)

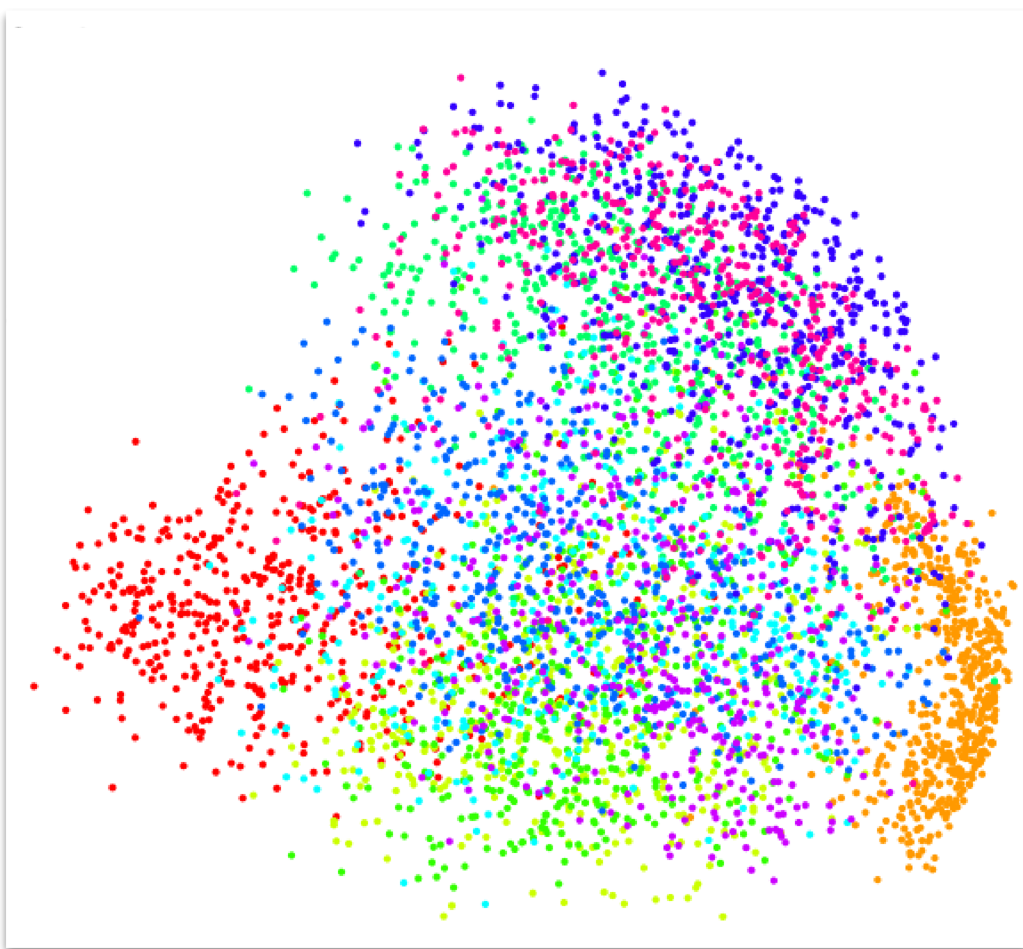


Dimensionality Reduction

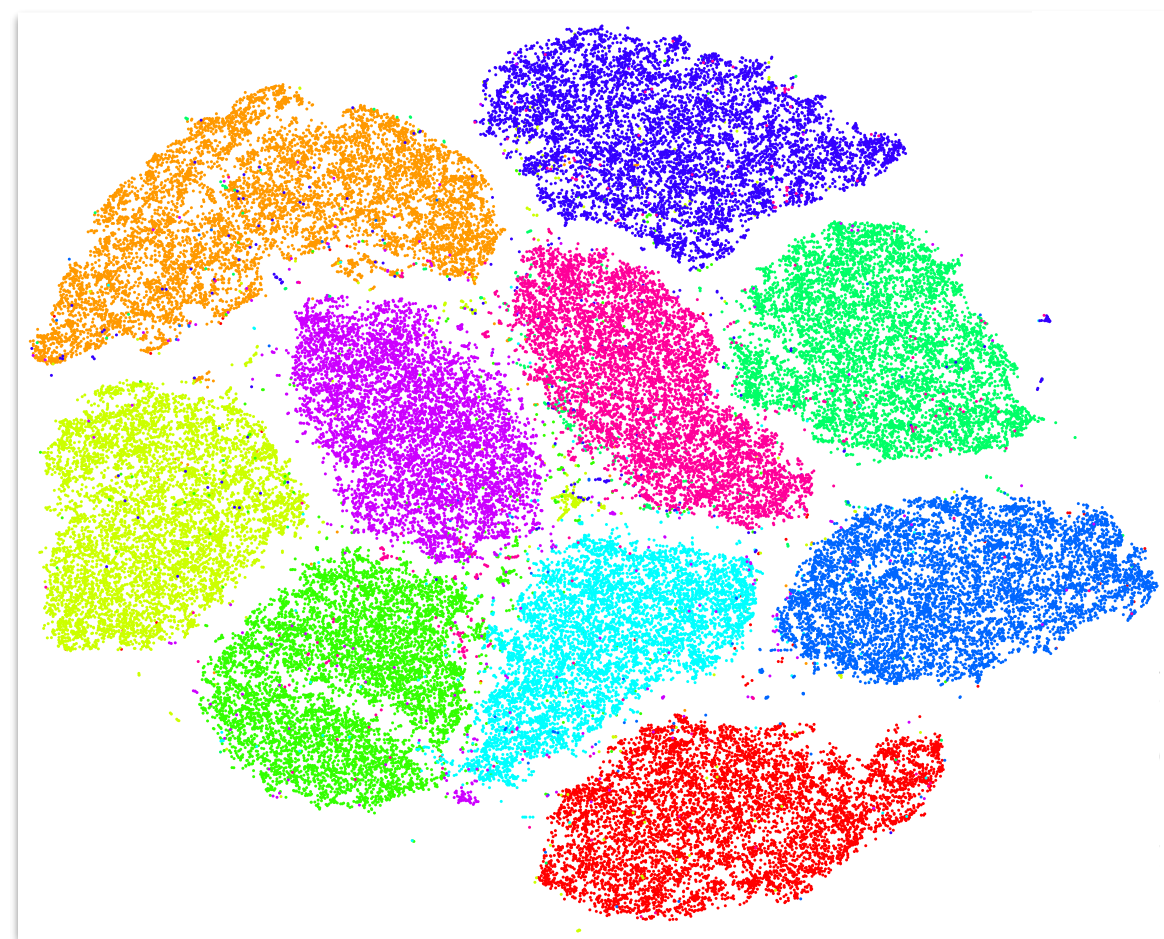
Goal: Map high dimensional data onto lower-dimensional data in a manner that preserves *distances/similarities*

Input Images

PCA (Linear)



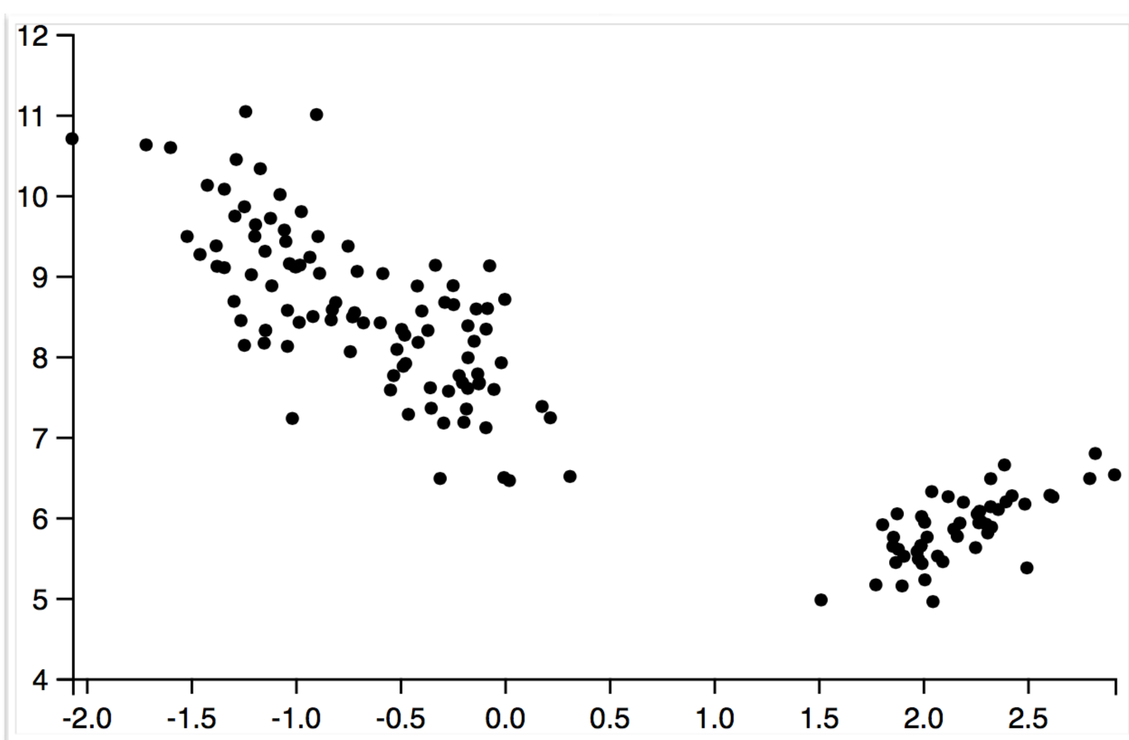
TSNE (Non-linear)



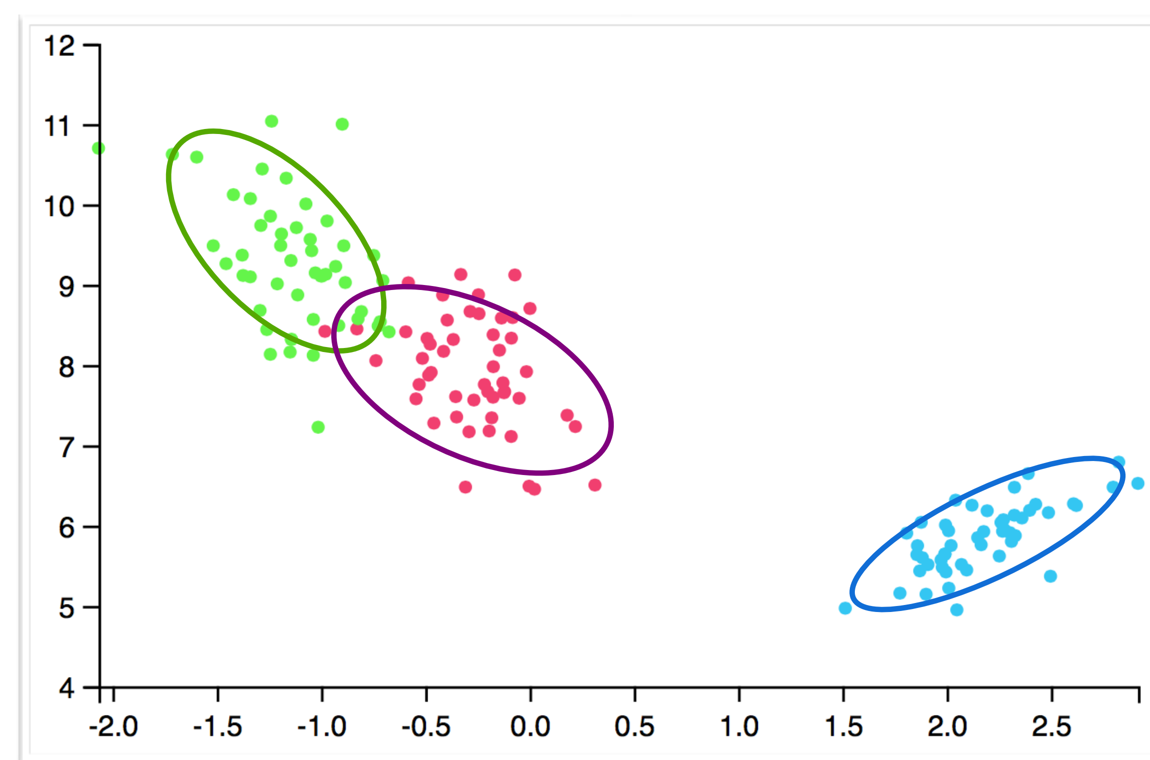
Clustering

Goal: Learn categories of examples
(i.e. classification without labels)

Iris Data (after PCA)



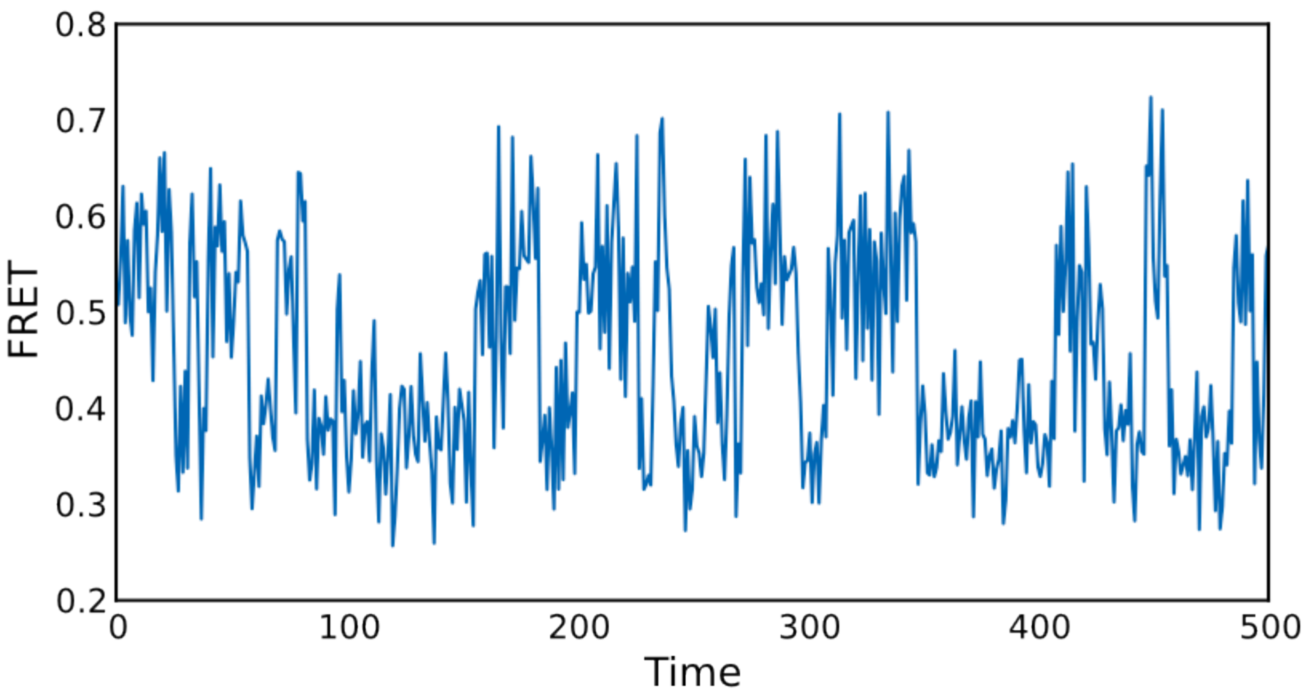
Inferred Clusters



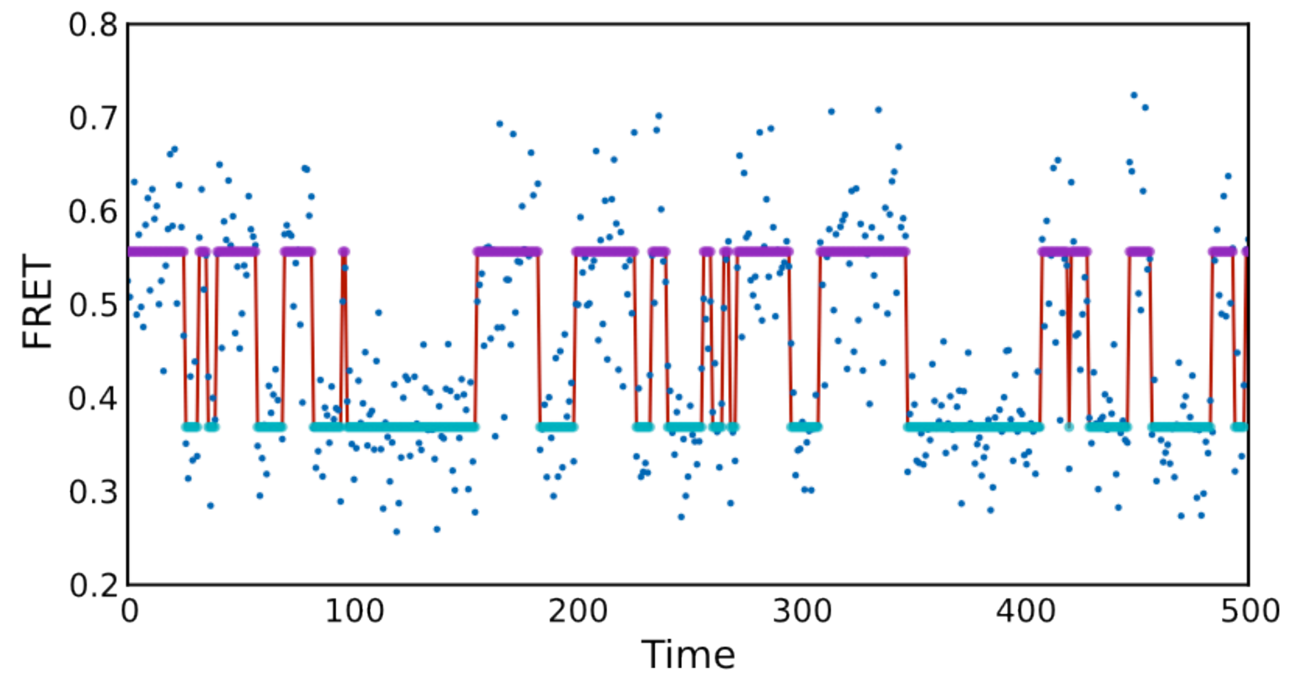
Hidden Markov Models

Goal: Learn categories of time points
(i.e. clustering of points within time series)

Time Series

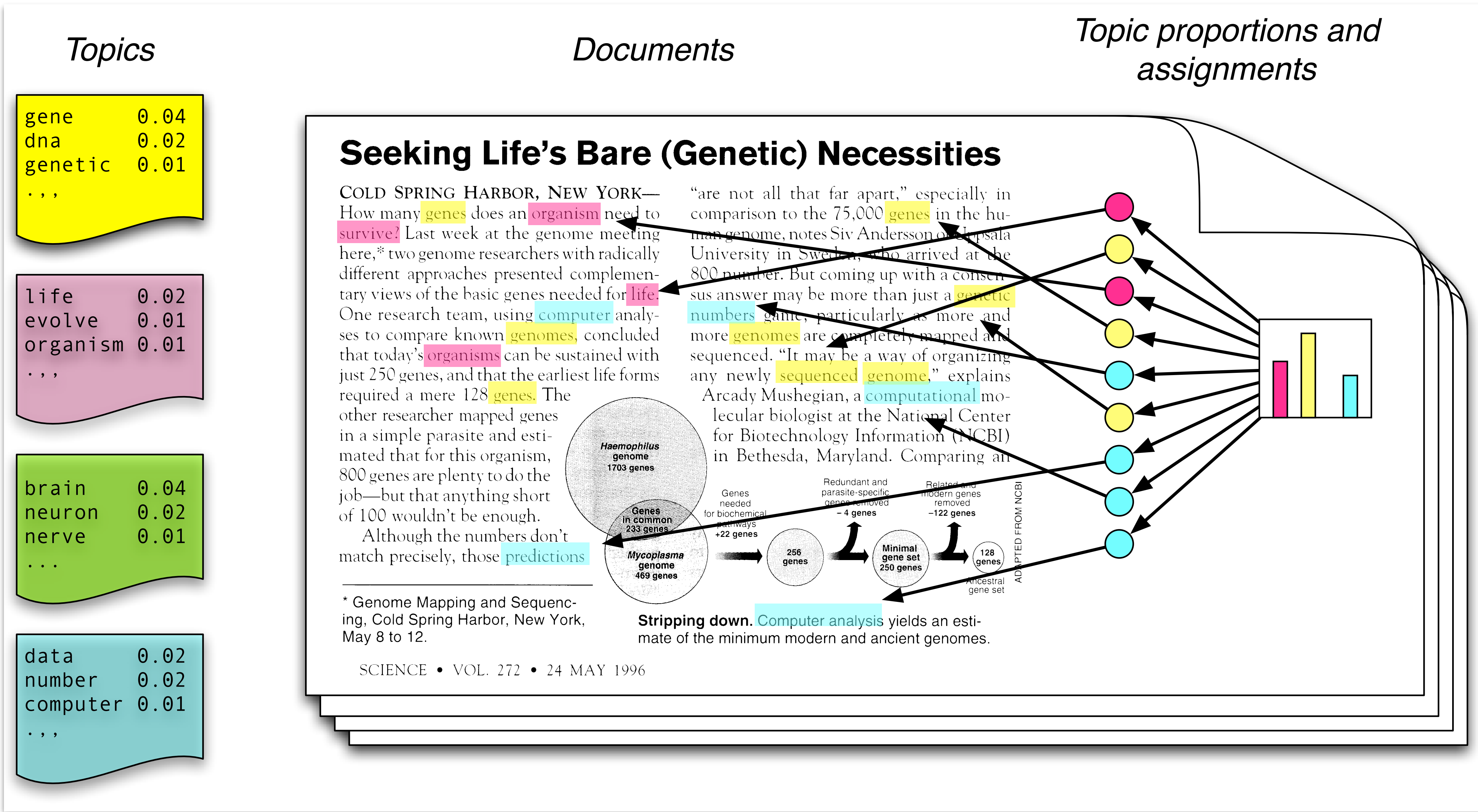


Sequence of States



Topic Models

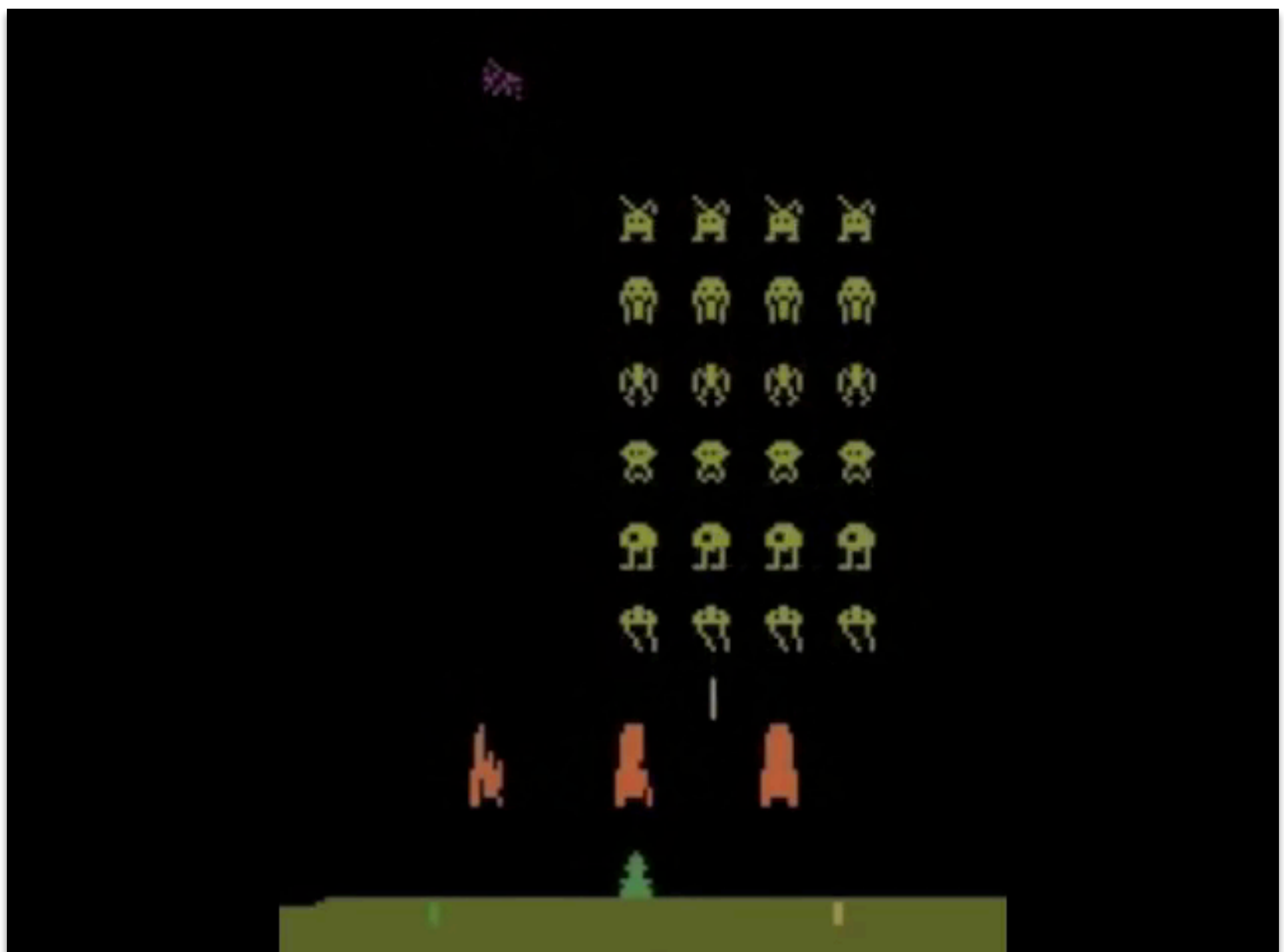
Goal: Learn topics (categories of words)
and quantify topic frequency for each document



Reinforcement Learning

Goal: Take *action* that maximizes future *reward*.

Example: Google Plays Atari

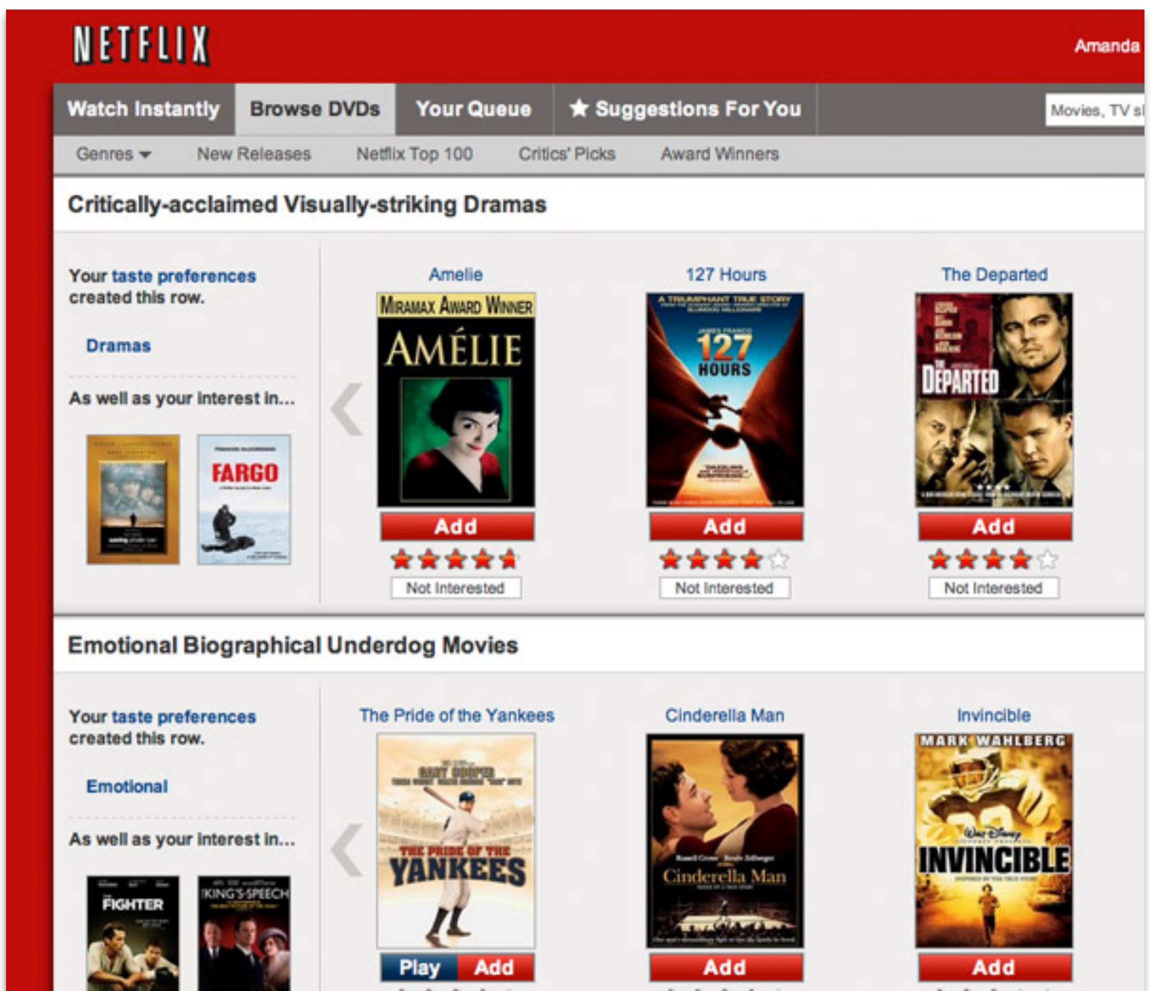


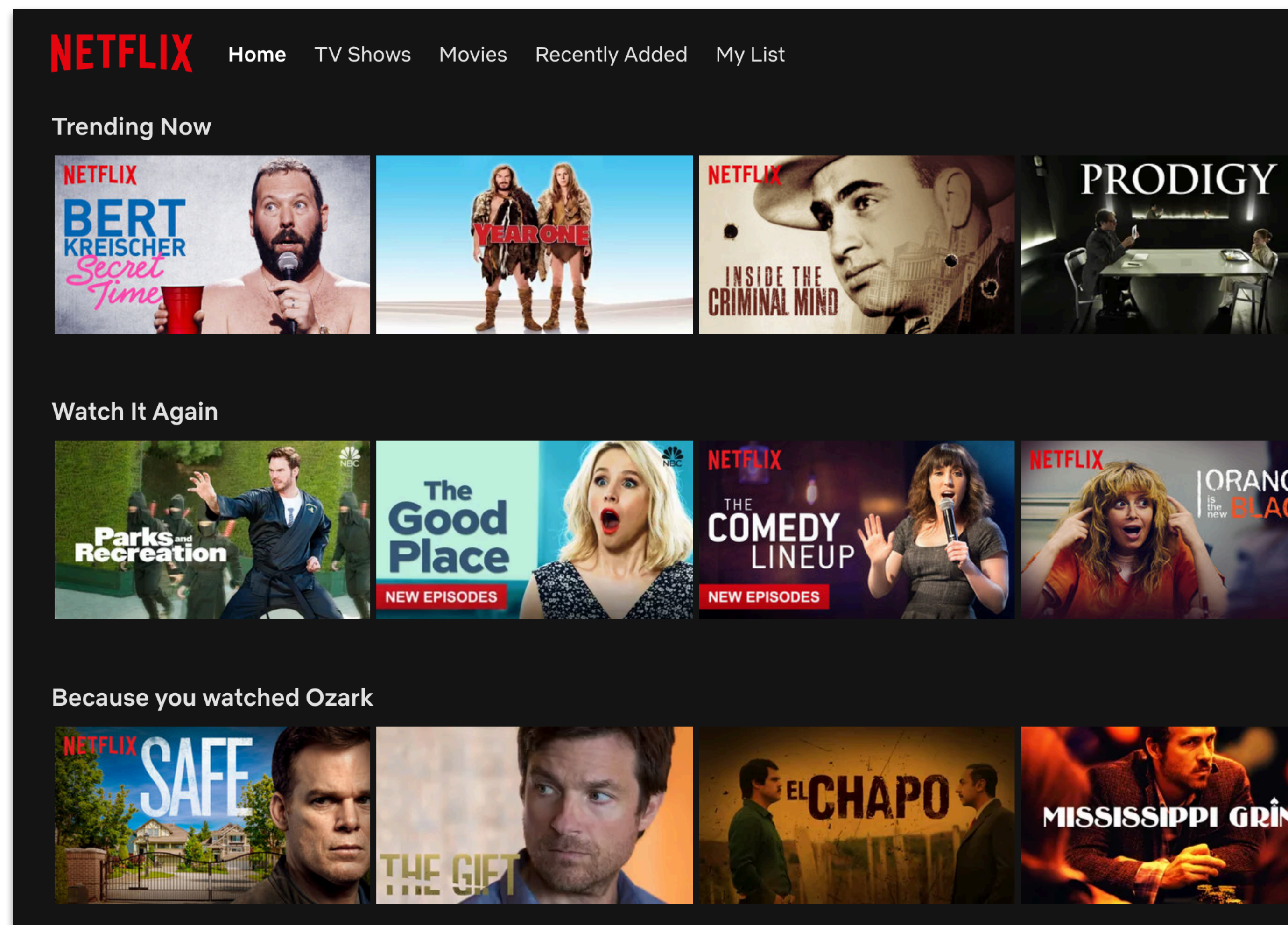
Action: Joystick direction / Buttons. **Reward:** Score.

Reinforcement Learning

Goal: Take *action* that maximizes future *reward*.

Example: Netflix Website Design





Action: Which movies to show. **Reward:** User Retention.

Theme: Optimization of Objectives

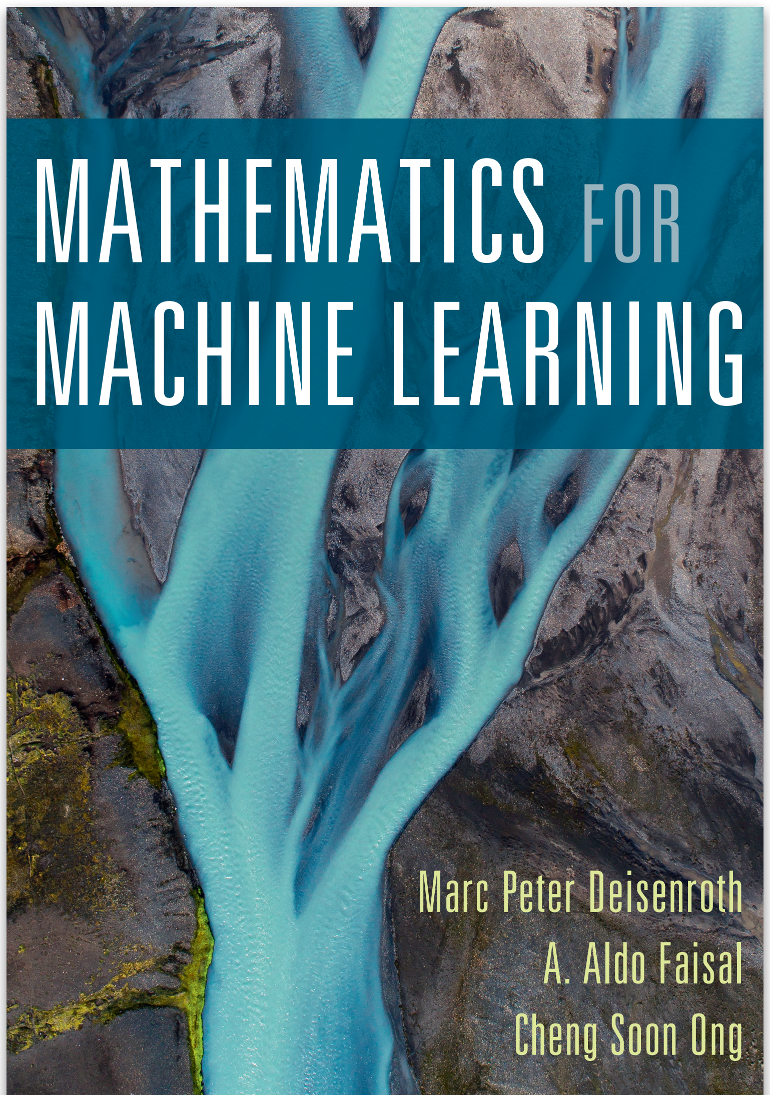
Common theme in Machine Learning:
Using data-driven algorithms to make *predictions* that are *optimal* according to some *objective*.

Supervised Learning: Minimize regression or classification loss

Unsupervised Learning: Maximize *expected* probability of data

Reinforcement Learning: Maximize *expected* reward

Materials



For first half-ish of class will use *Mathematics for Machine Learning*; this is **free** and online.

Other readings are also from free, online sources (CIML and Elements of Statistical Learning)

https://mml-book.com/

Grading

30%	HOMEWORKS
5%	IN CLASS EXERCISES
30%	MID-TERM
35%	FINAL PROJECT

Project

Teams of 1-2 — bigger gets unwieldy

- Obviously, I expect more from teams of 2 than from 1

Select a problem / dataset that is *interesting to you*

- Sarthak and I can help identifying projects if you're stuff
- In any case, talk to me about any ideas (early is good!)

Submit a report and present project

Project Deadlines (*Tentative*)

- **3/13** Submit a project description — < 1 page
- **3/17** Project pitches in class / feedback; this should include exploratory data analysis (or some sort of prelim results)
- **4/7** and **4/9** project presentations
- **4/14** final reports due

Questions for me??