DASIL Python Workshop # 4

October 13th 2022

Introduction to Machine Learning Part #3

Yusen He, PhD - DASIL Data Scientist

Prof. Julia Bauder – Director of DASIL

Martin Pollack – DASIL Post Bachelor Fellow

DATA ANALYSIS & SOCIAL INQUIRY LAB

GRINNELL COLLEGE

Intro to Machine Learning Part #3

AGENDA

Other Types of Neural Networks

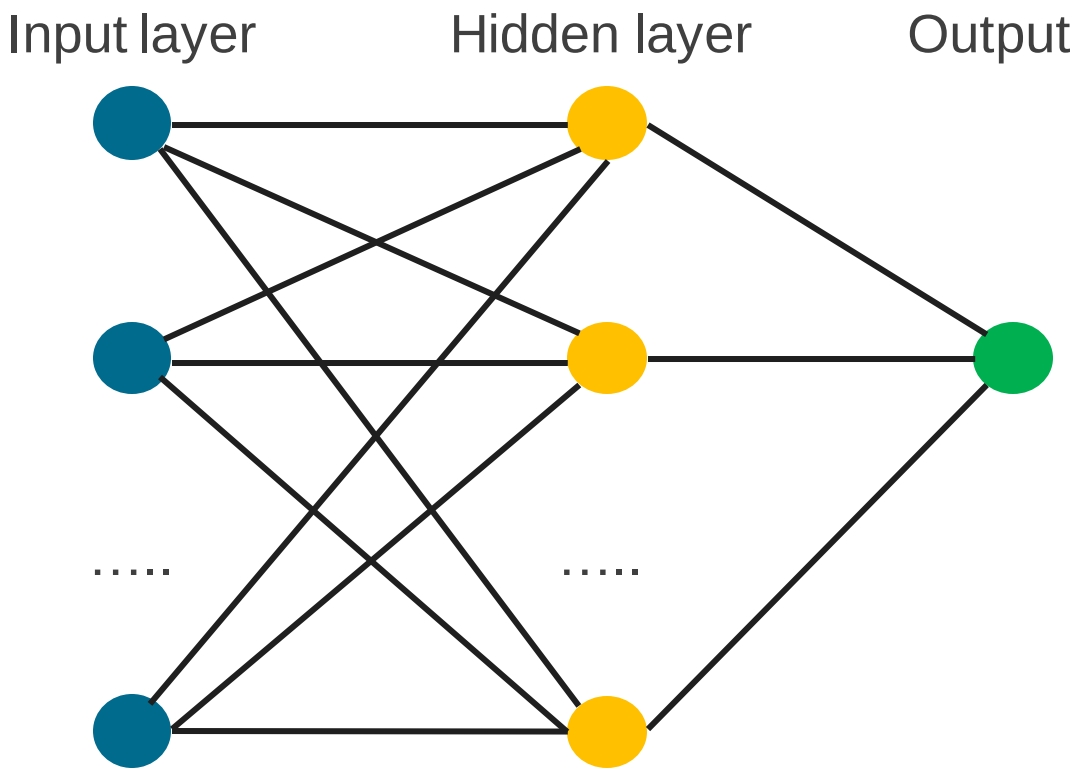
Model Evaluation

Unsupervised Learning Basics

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

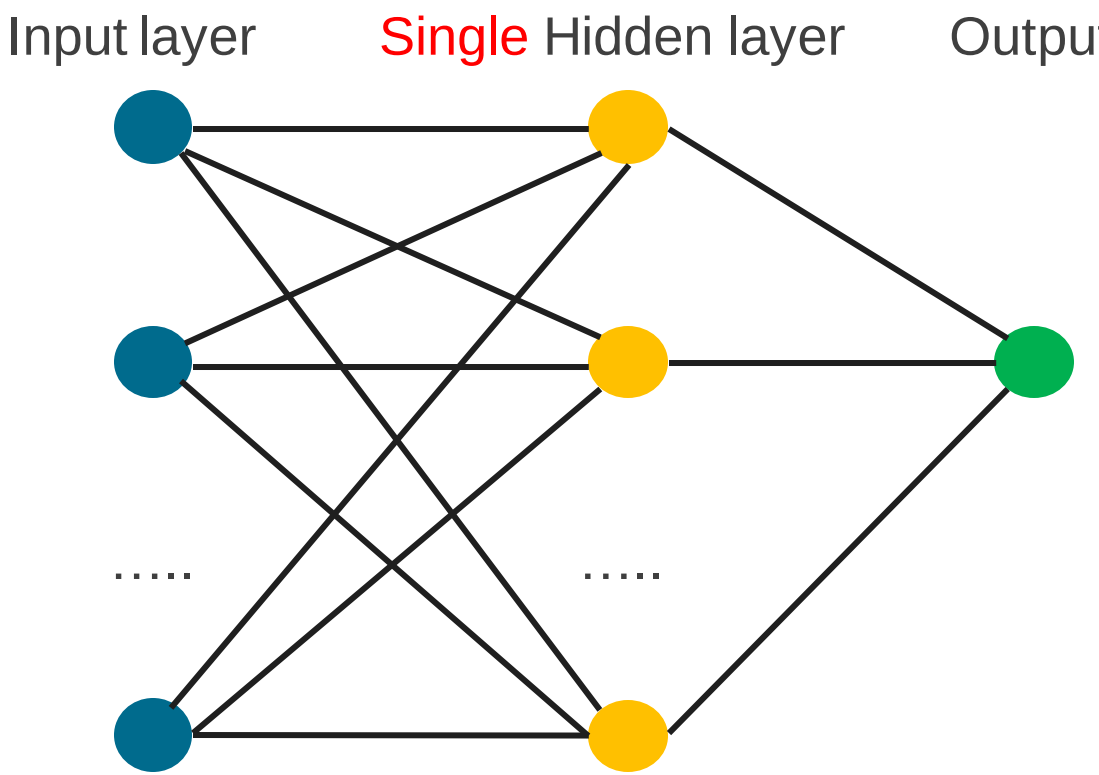
- Recall: Artificial Neural Network



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

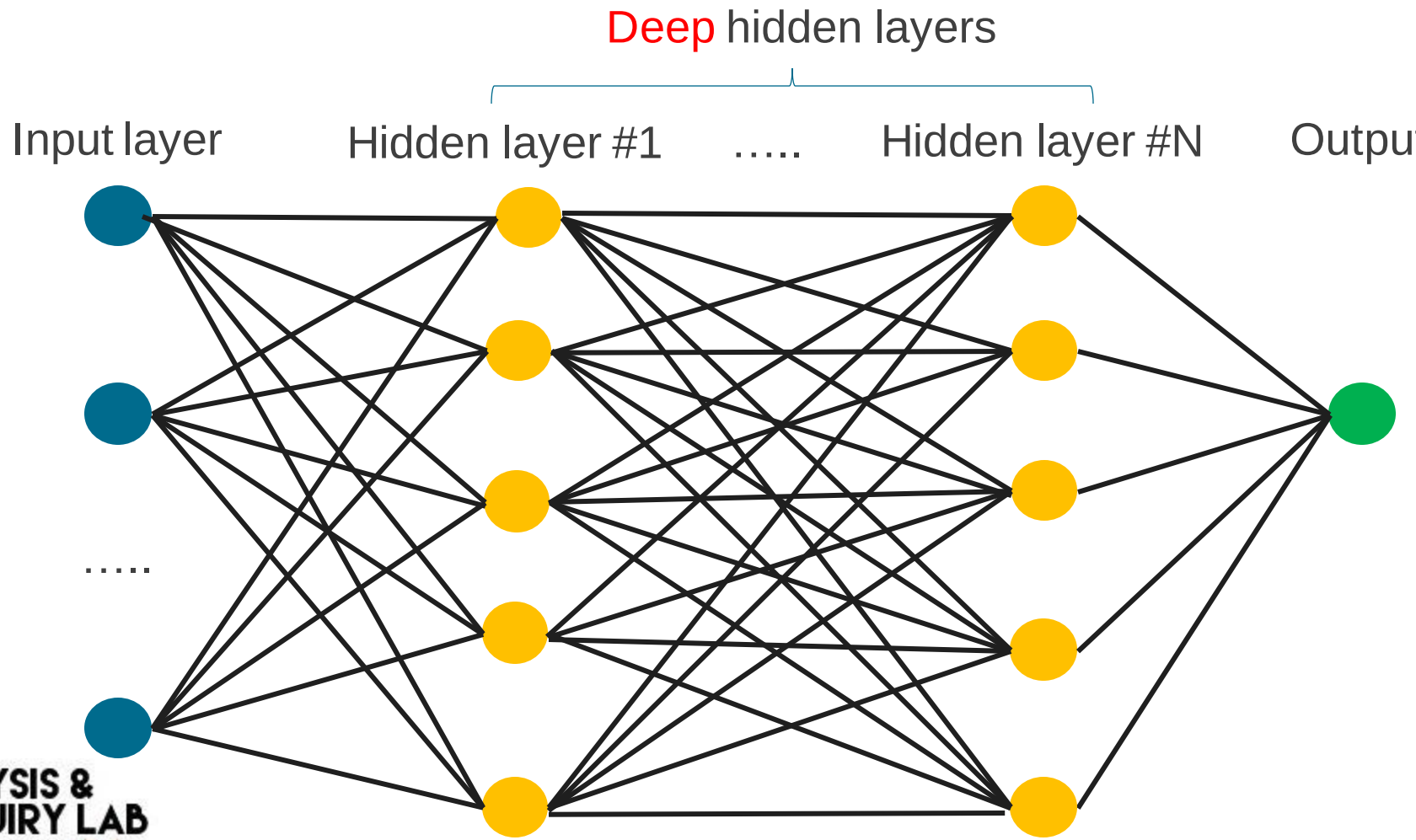
- Shallow Neural Network



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

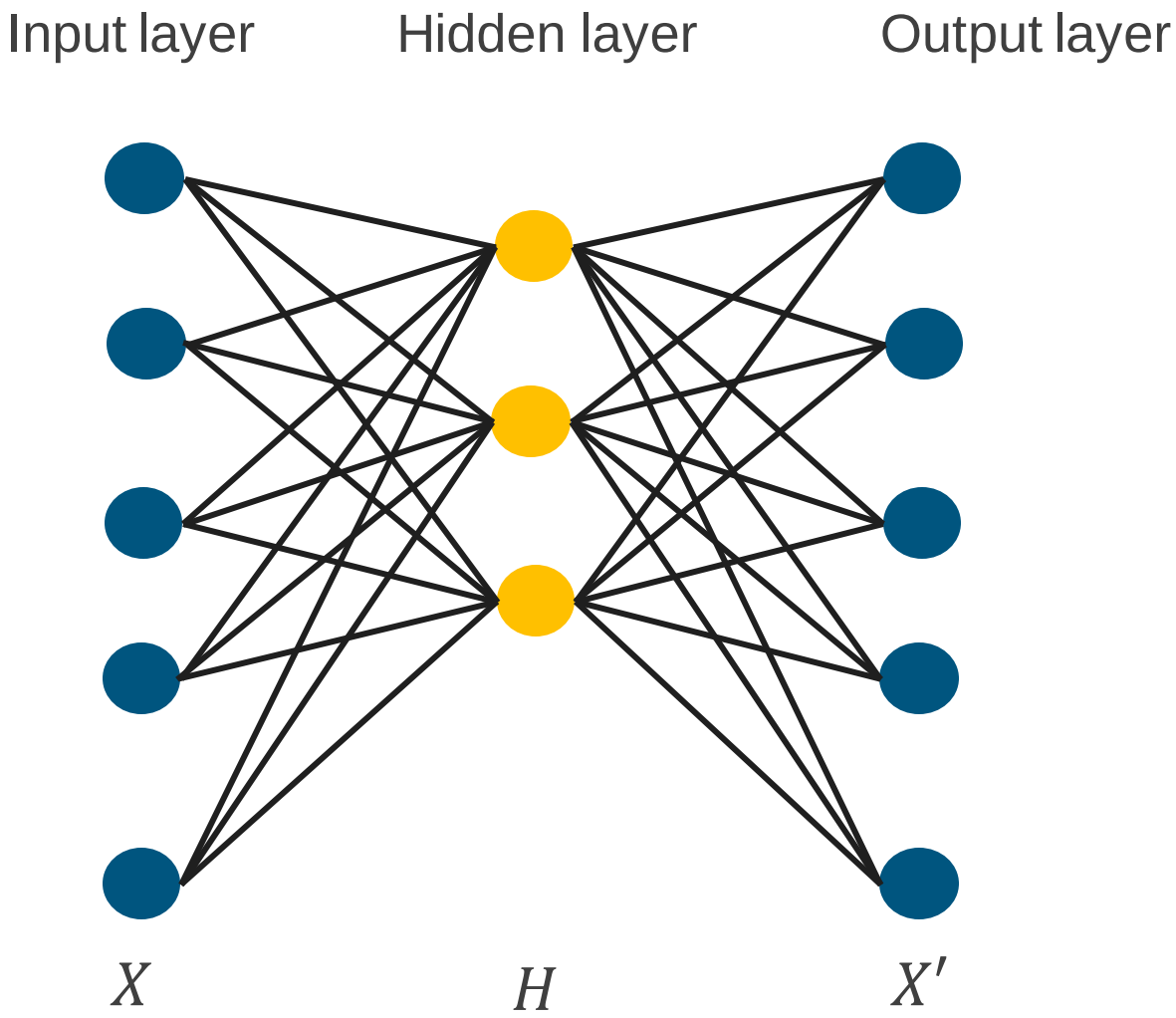
- Deep Neural Network



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

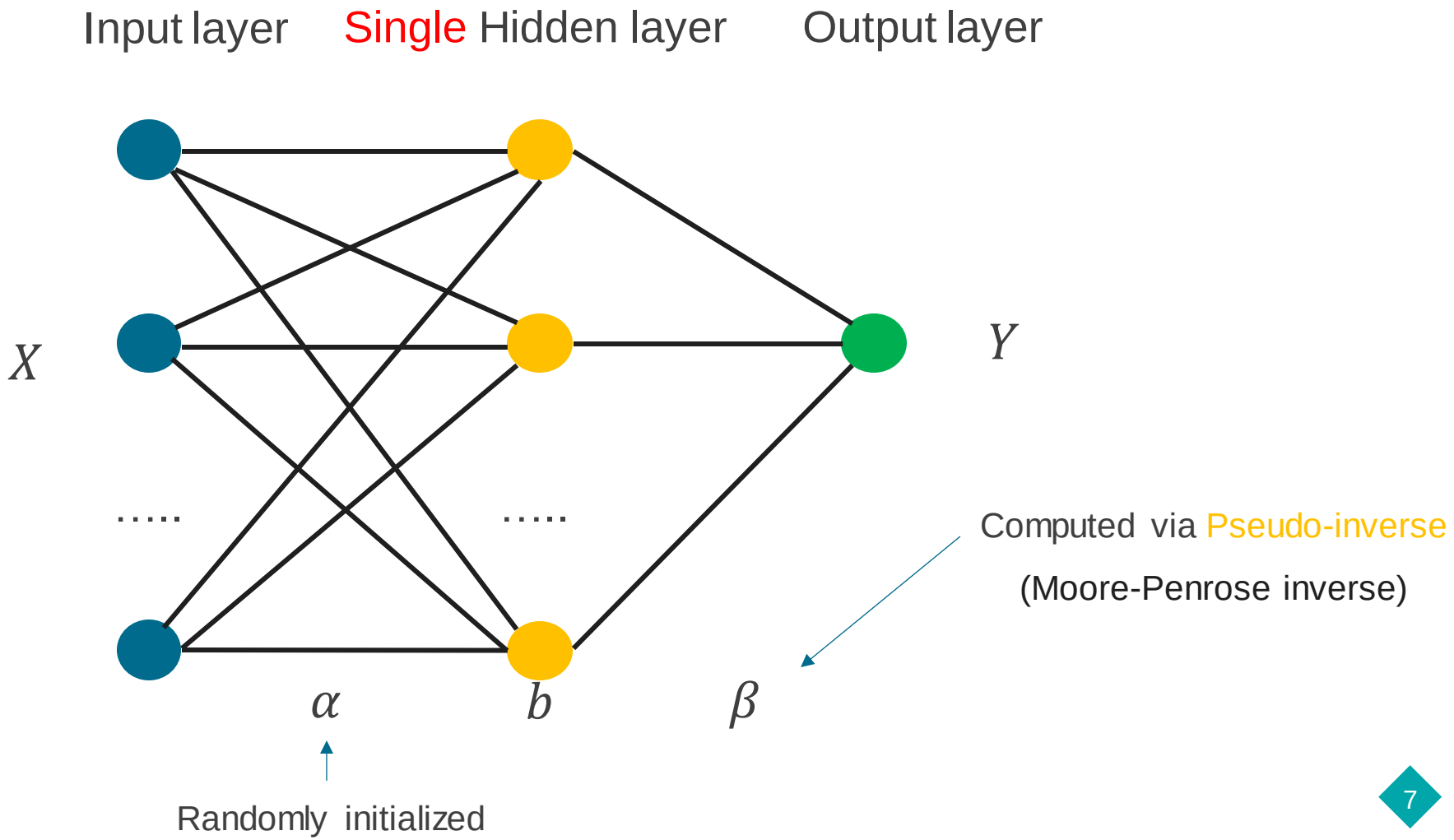
- Auto-Encoder



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

- Extreme Learning Machine (ELM) (Huang et al. 2006)



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

- Convolutional Neural Networks

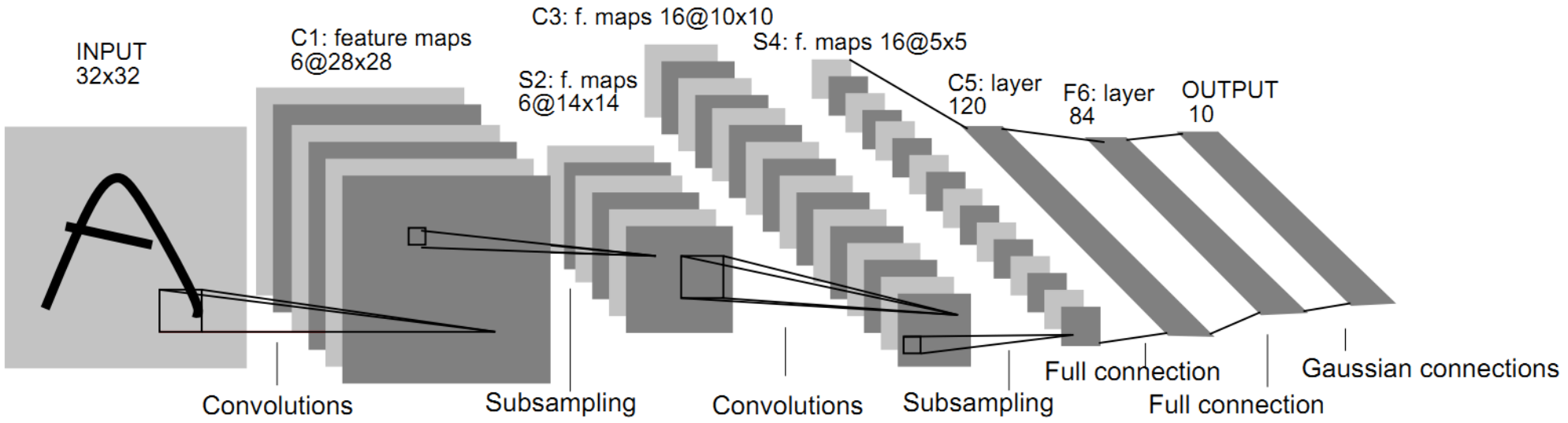


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

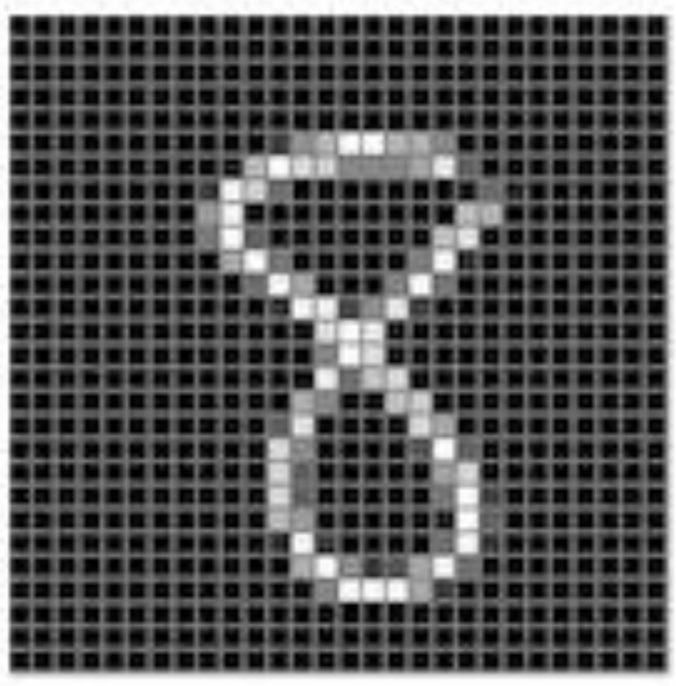
LeCun, Bottou, Bengio, & Haffner. (1998)

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

- Convolution Operation (Dot products):

What you see:



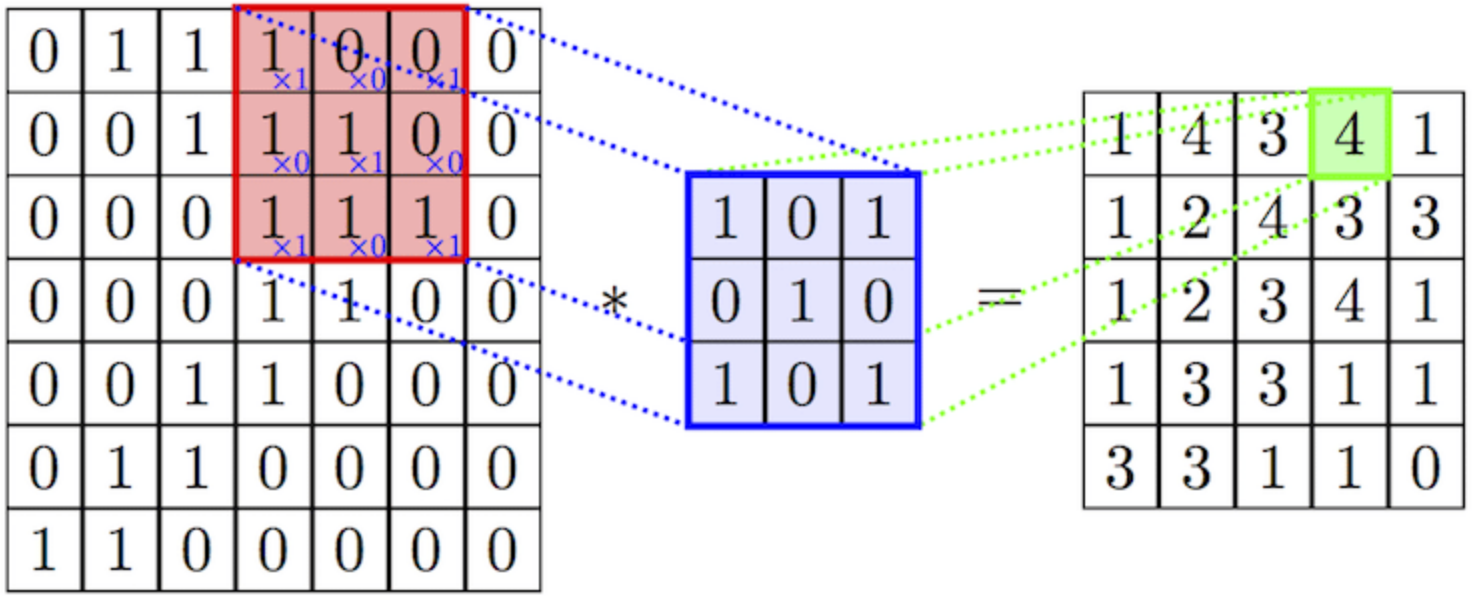
What computer sees:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	4	62	146	182	254	254	181	176	139	15	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	34	186	253	217	208	136	136	136	166	232	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	61	242	208	111	3	0	0	0	0	0	18	32	107	43	0	0	0	0	0	0
0	0	0	0	0	0	0	0	156	242	23	0	0	0	0	0	0	0	13	191	181	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	121	255	98	3	0	0	0	0	0	8	194	225	12	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	169	253	120	3	0	0	0	0	128	247	51	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	3	111	244	169	19	0	14	131	249	117	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	59	241	235	72	142	229	66	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	25	218	254	231	36	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	133	253	221	33	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	19	237	111	196	217	19	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	138	0	23	193	204	18	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	96	224	0	0	0	25	218	169	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	138	0	0	0	0	86	253	99	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	3	162	214	11	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	215	97	0	0	0	0	0	118	253	68	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	185	157	0	0	0	0	0	40	254	98	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	50	244	61	0	0	0	0	112	244	58	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	174	251	142	59	83	167	244	111	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	6	133	253	253	253	169	61	3	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

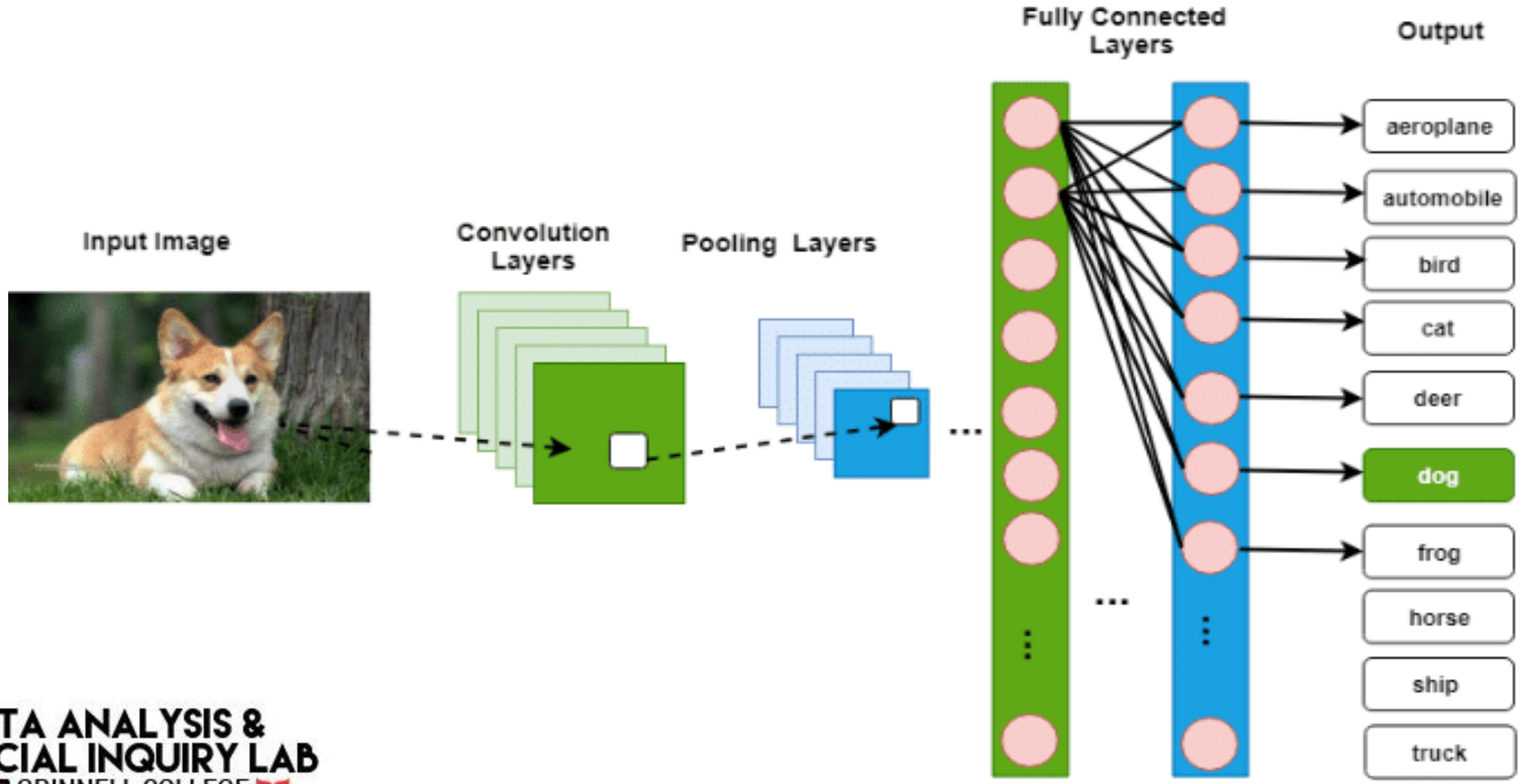
- Convolution Operation (Dot products):



Input image Convolution filter (kernel) Activation map

Other Neural Networks

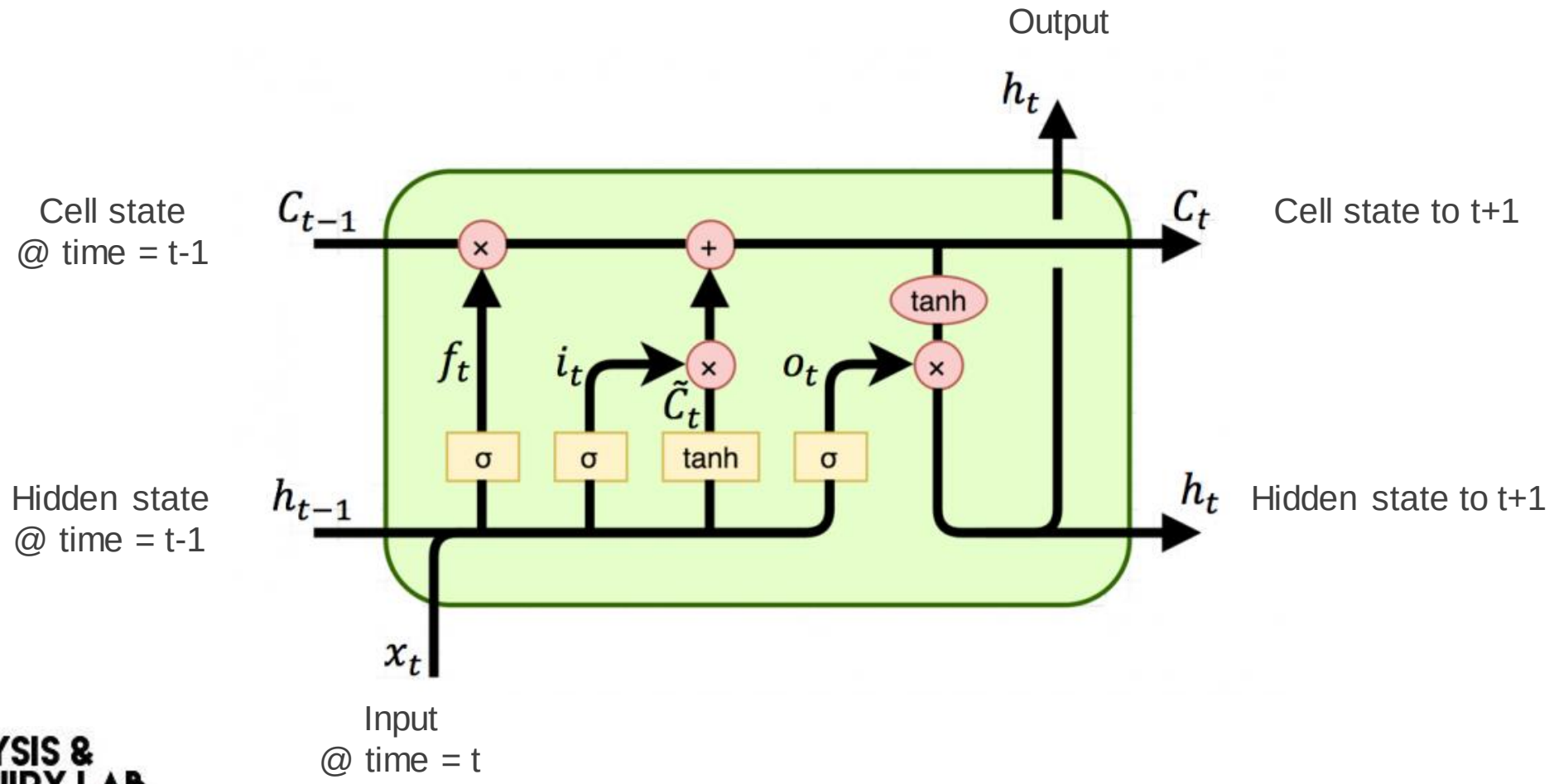
- Typical CNN Structure:



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

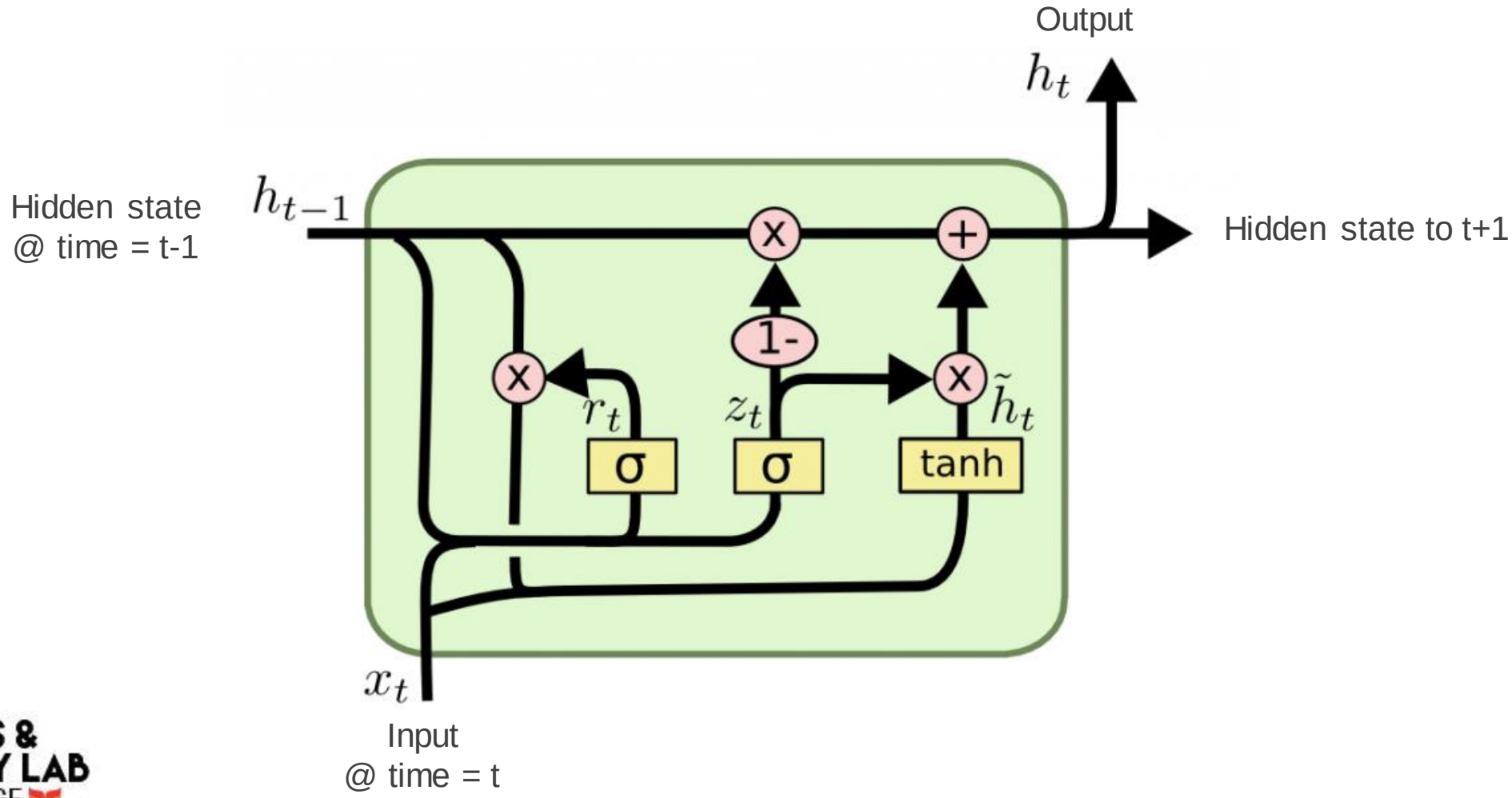
- Long-Short-Term-Memory (LSTM)



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Other Neural Networks

- Gated Recurrent Unit (GRU)



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Intro to Machine Learning Part #3

AGENDA

Other Types of Neural Networks

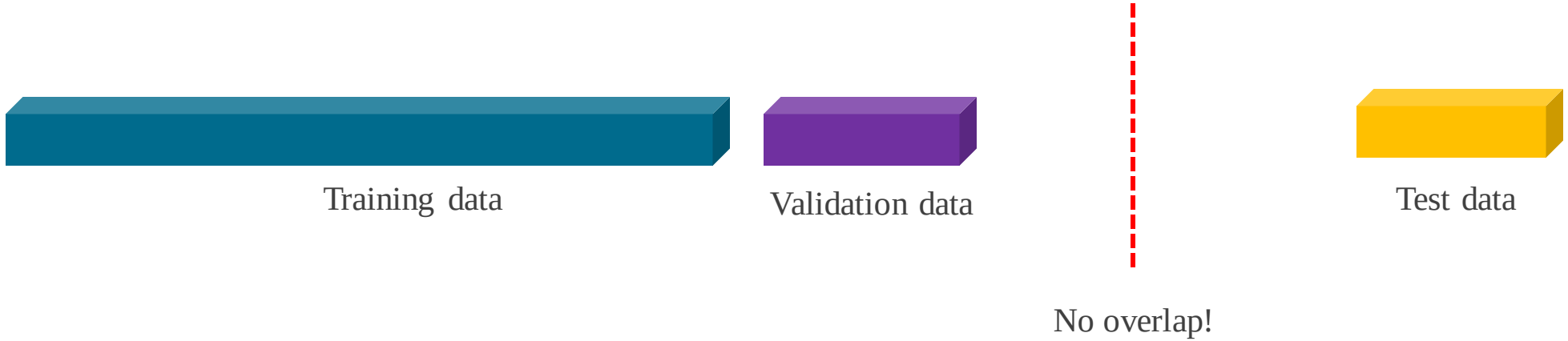
Model Evaluation

Unsupervised Learning Basics

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Model Evaluation

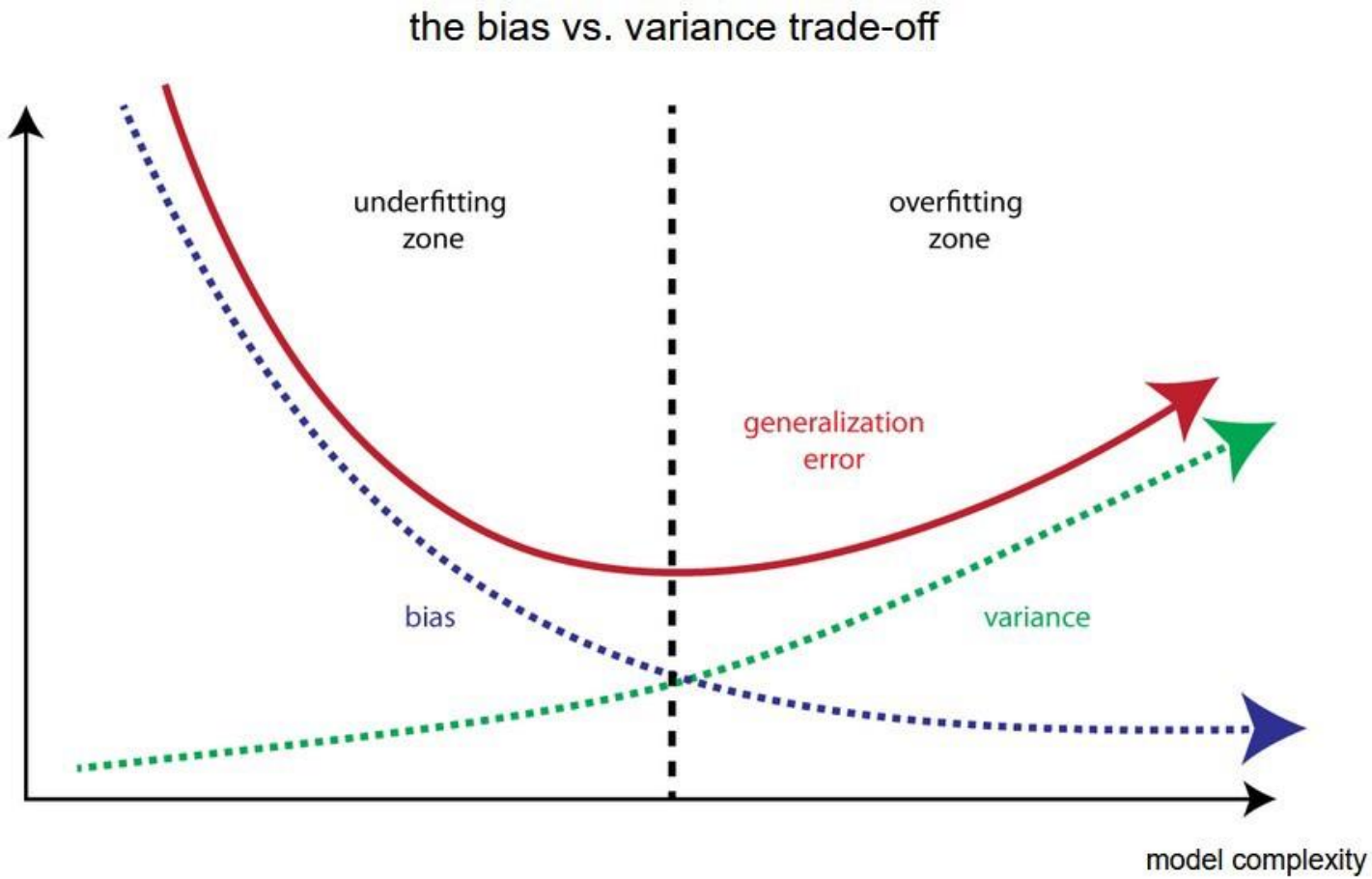
Recall: Training & Validation Dataset



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Model Evaluation

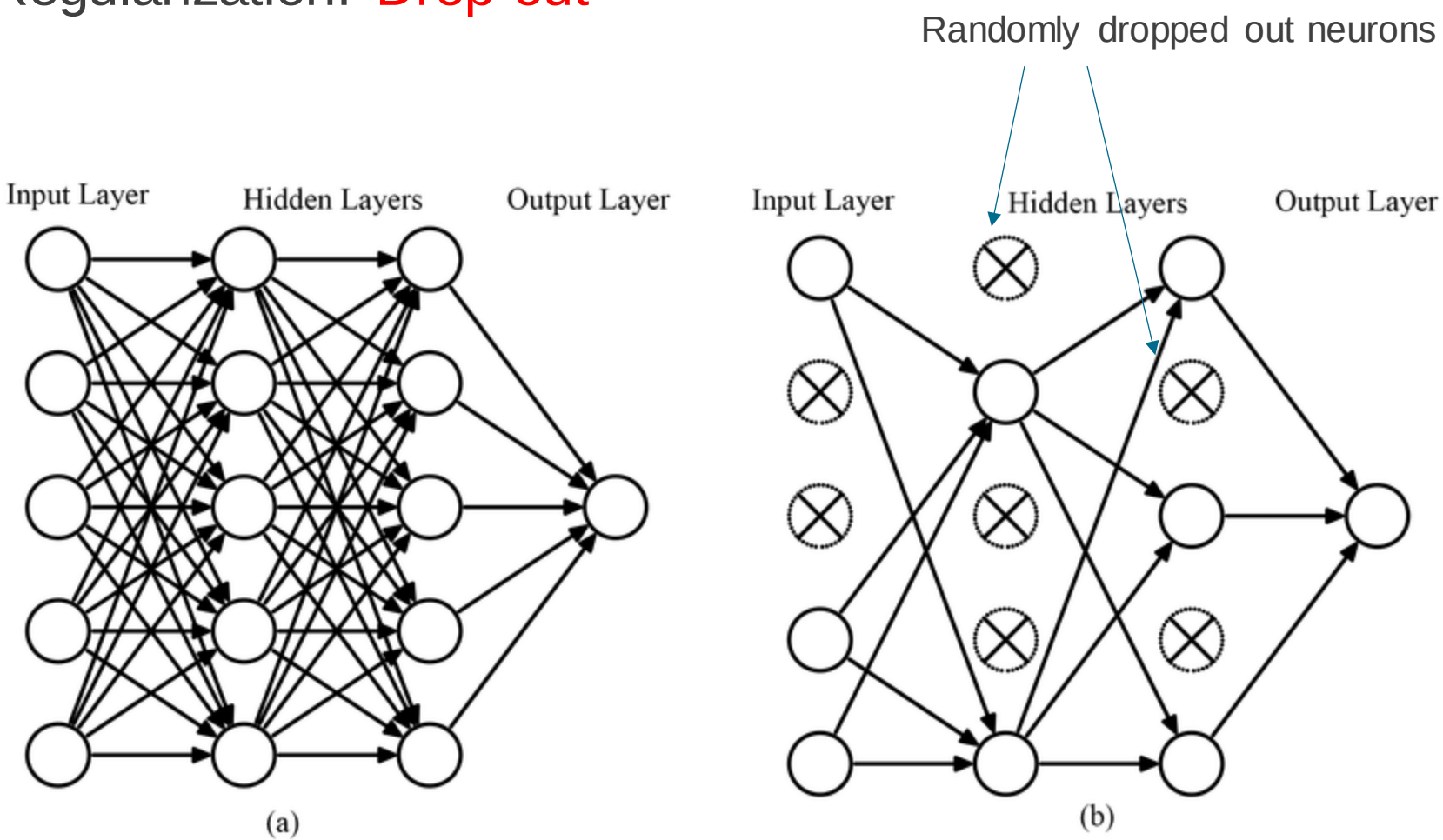
Bias Variance Trade Off:



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Model Evaluation

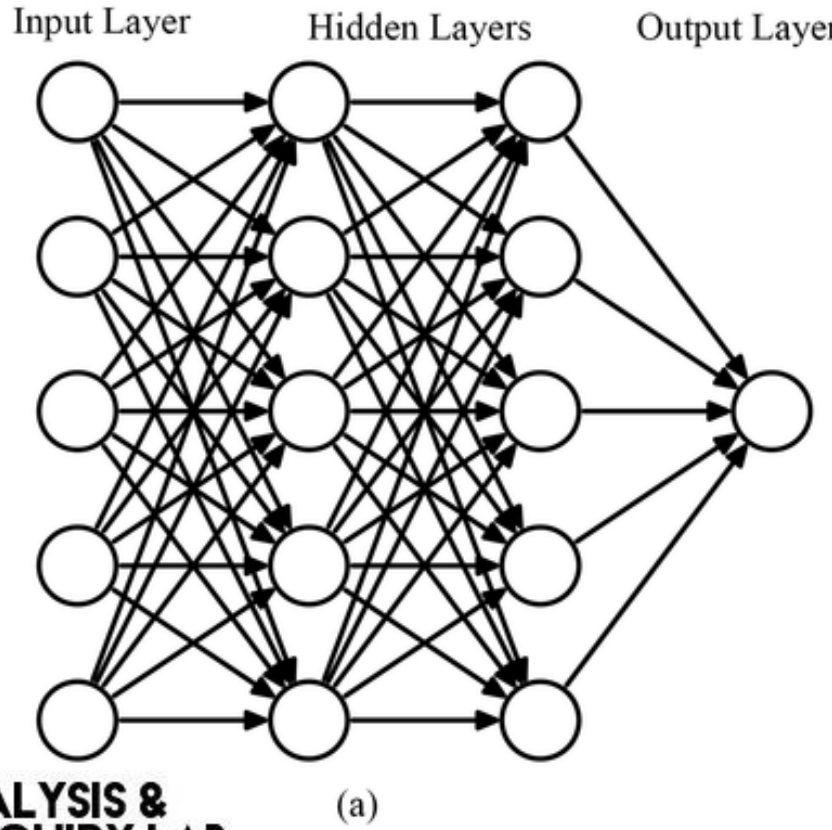
Neural Network Regularization: Drop out



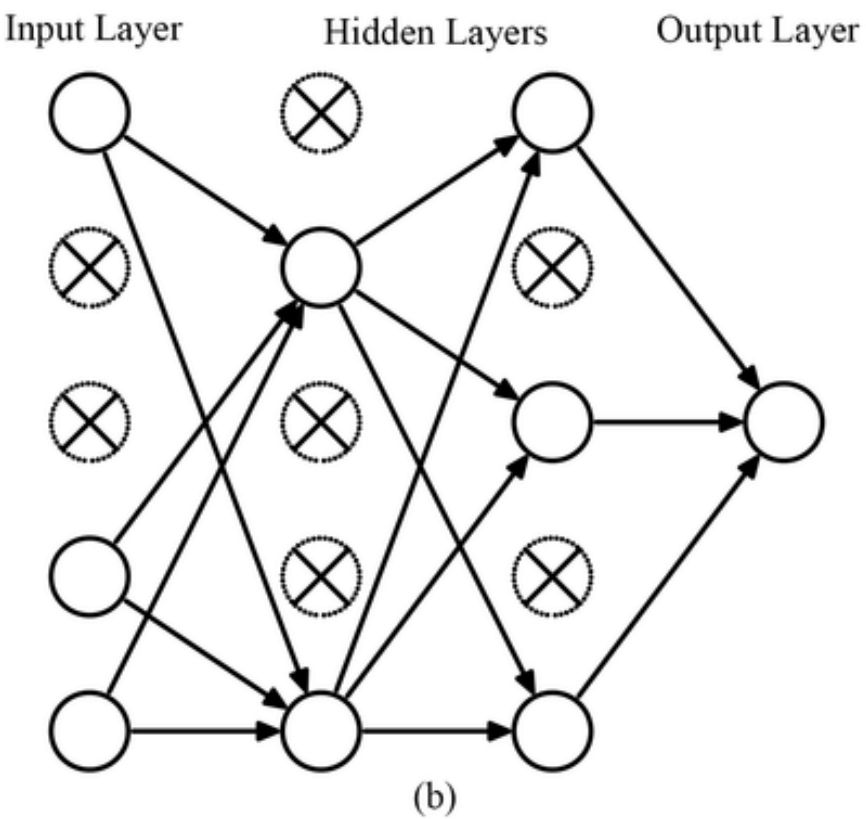
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Model Evaluation

Discussion: Which neural network is better?



(a)



(b)

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Intro to Machine Learning Part #3

AGENDA

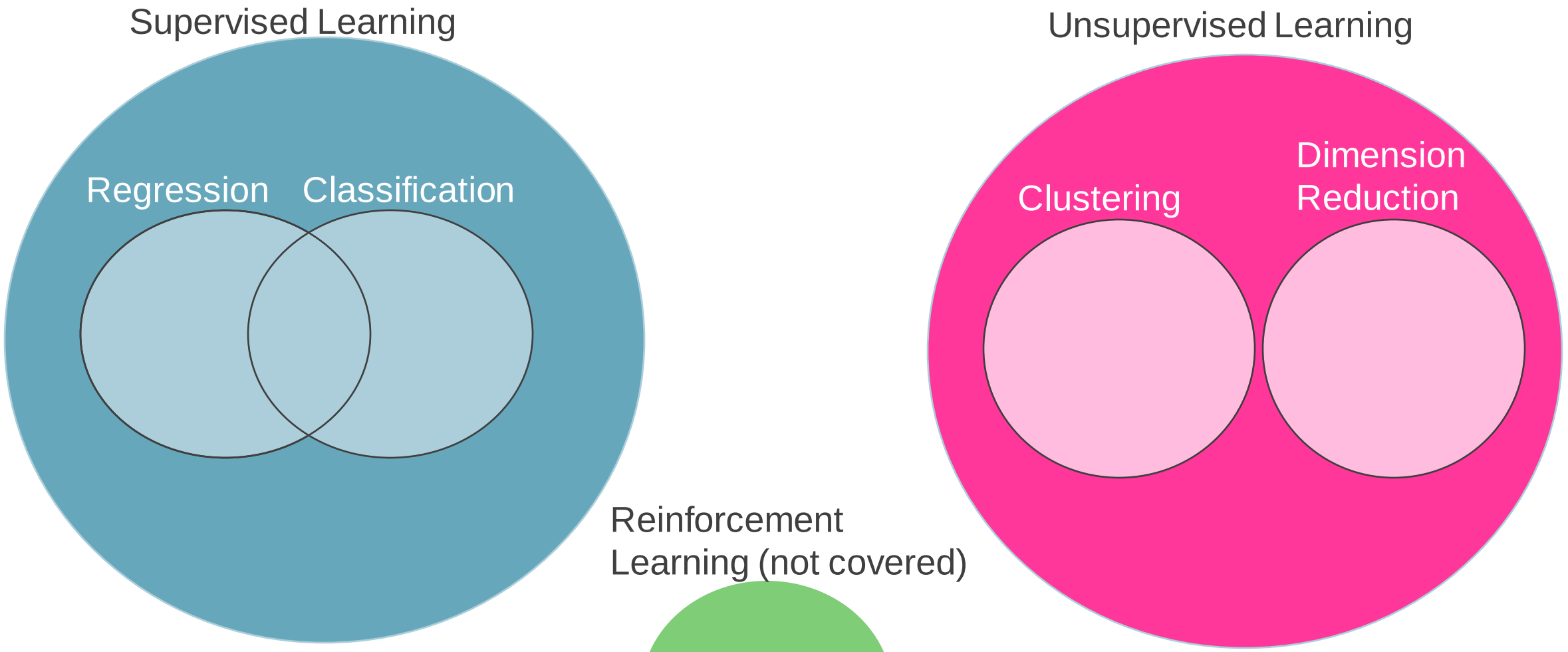
Other Types of Neural Networks

Model Evaluation

Unsupervised Learning Basics

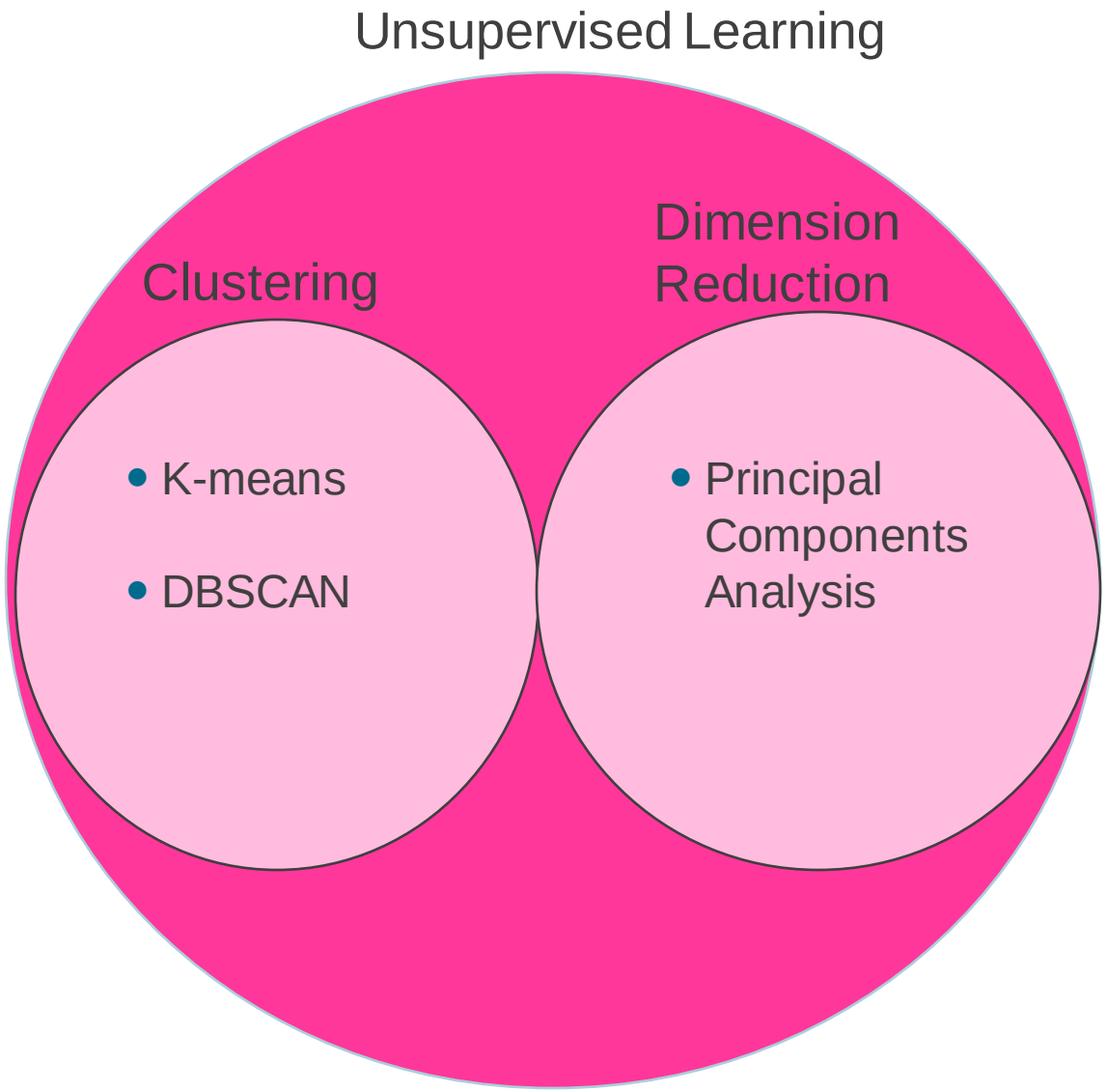
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

What is Machine Learning?



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Algorithms - Unsupervised Learning



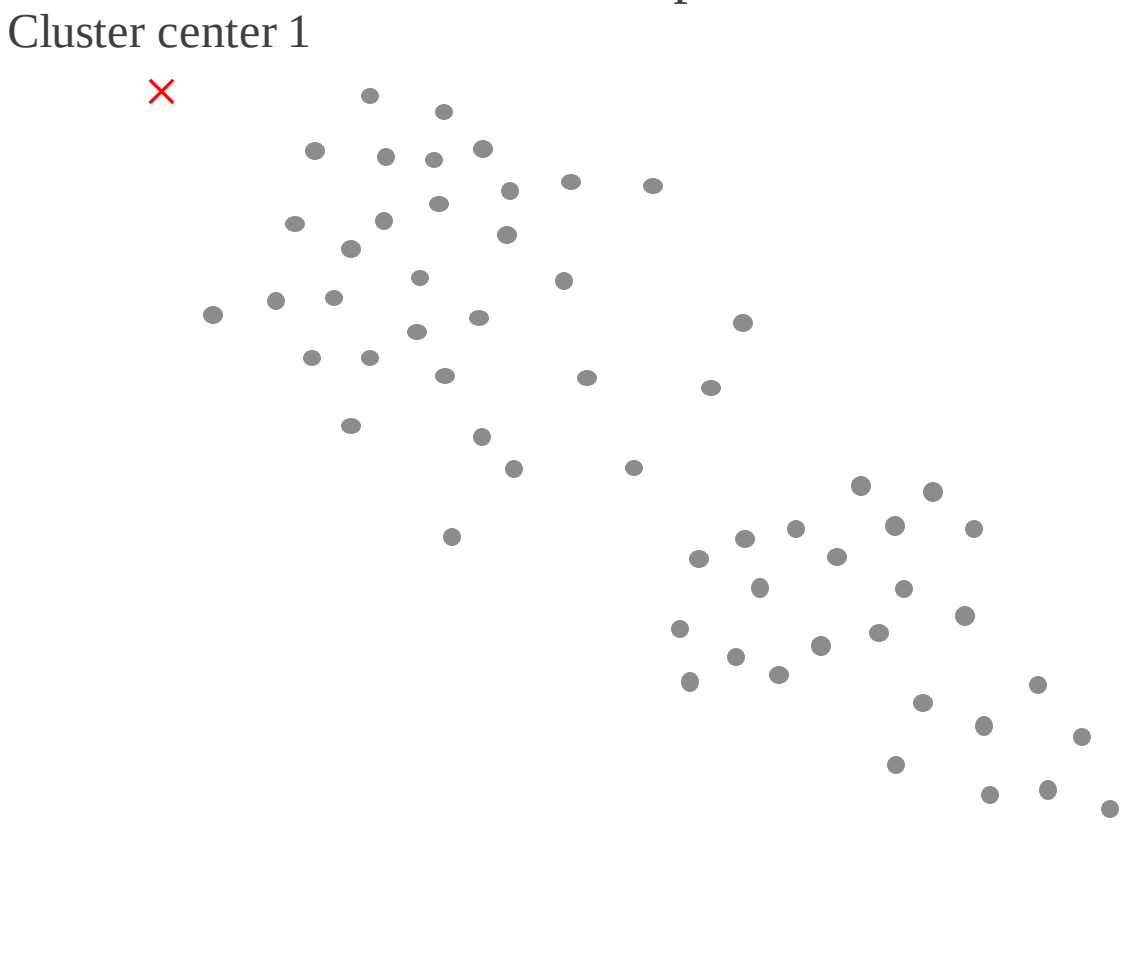
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning - Clustering

- K-Mean Clustering

  Lloyd's algorithm

Step #1: Random Mean Initialization



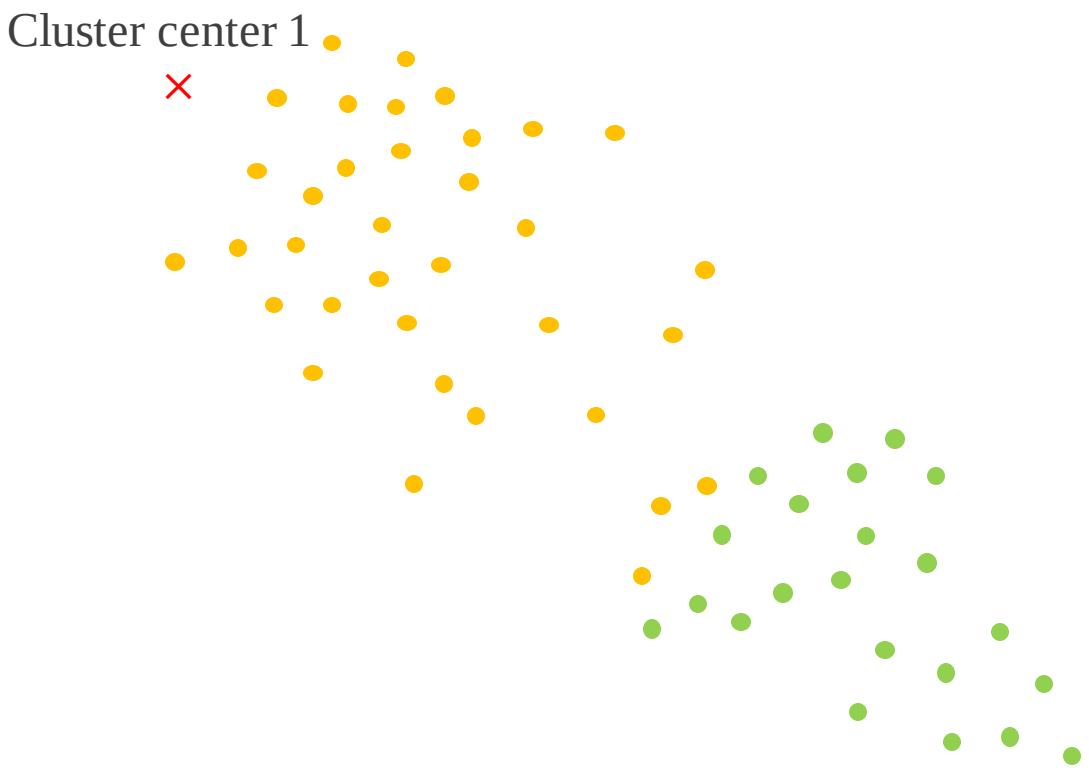
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning - Clustering

- K-Mean Clustering:

  Lloyd's algorithm

Step #2: Assign class labels by distances



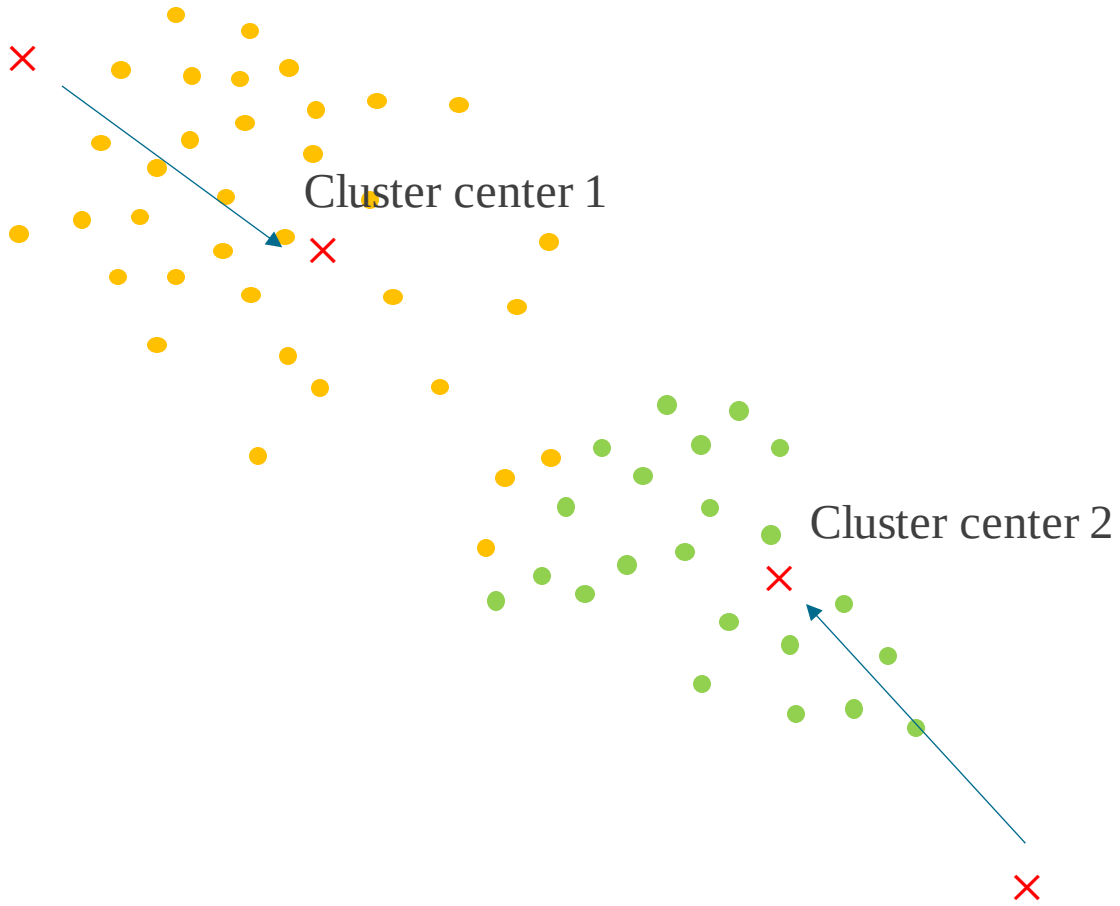
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning - Clustering

- K-Mean Clustering

  Lloyd's algorithm

Step #3: Calculate cluster centers



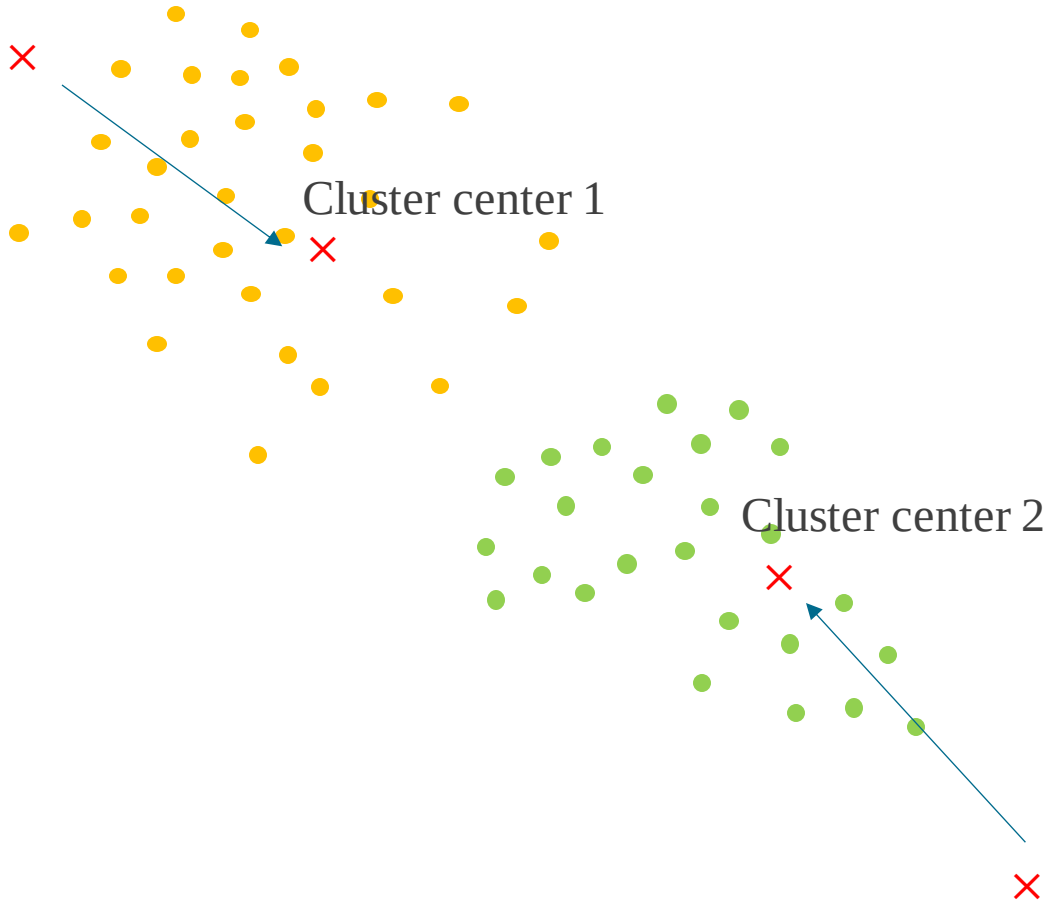
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning - Clustering

- K-Mean Clustering

  Lloyd's algorithm

Step #4: Re-assign labels by distance



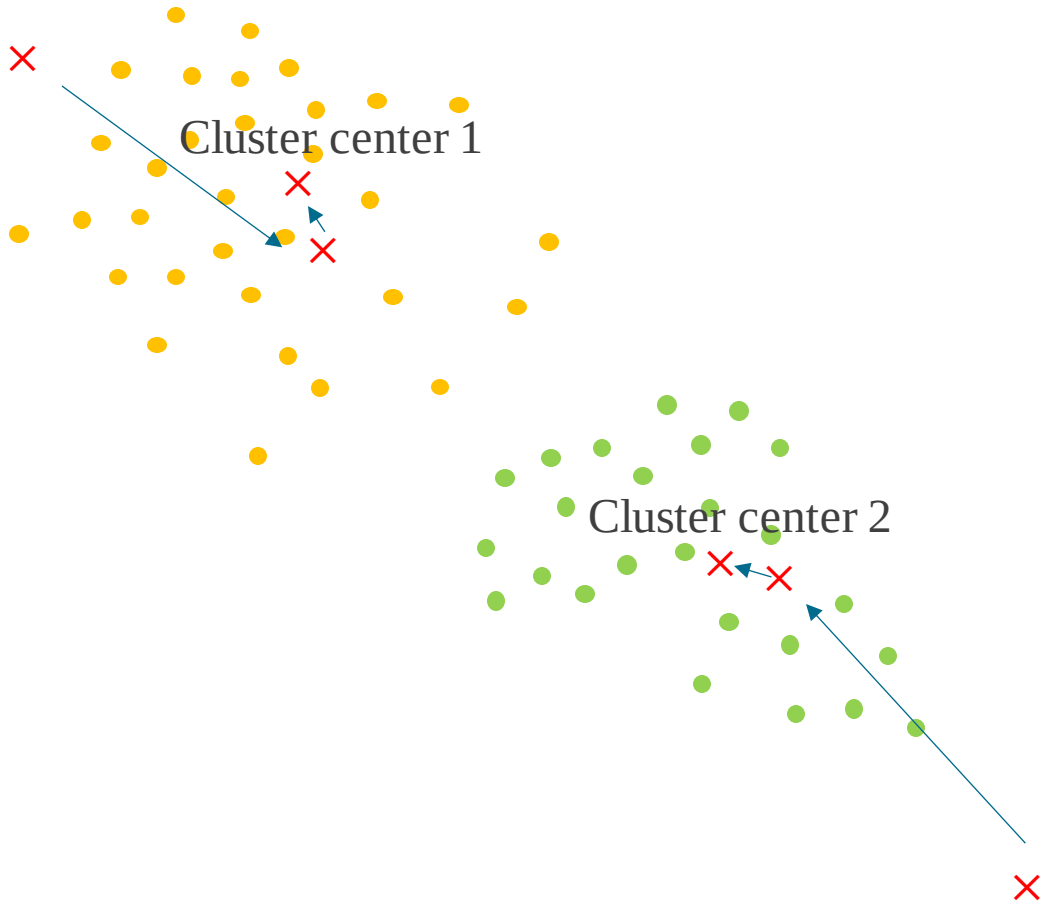
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning - Clustering

- K-Mean Clustering

  Lloyd's algorithm

Step #5: Re-calculate cluster centers

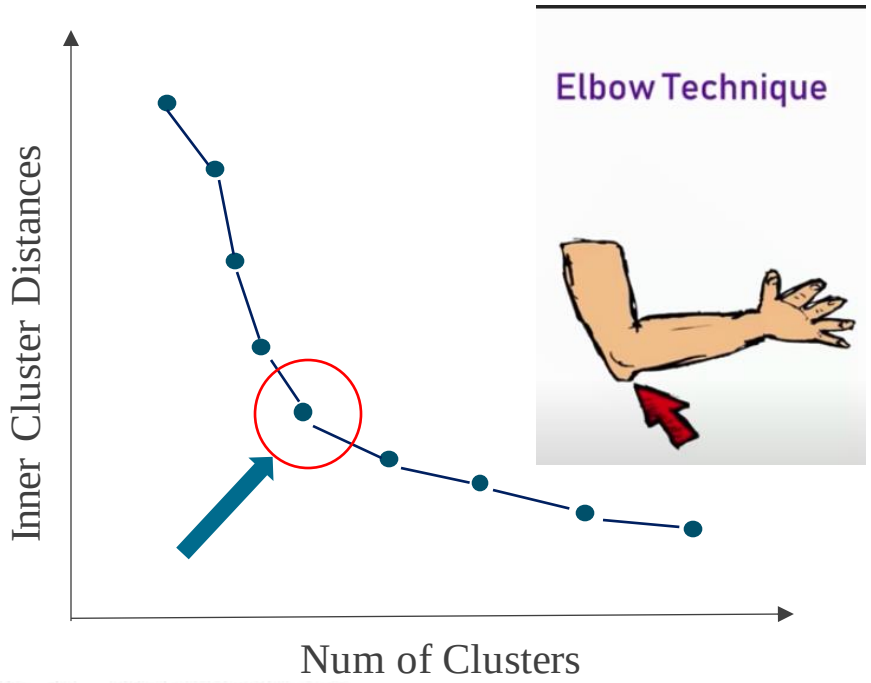


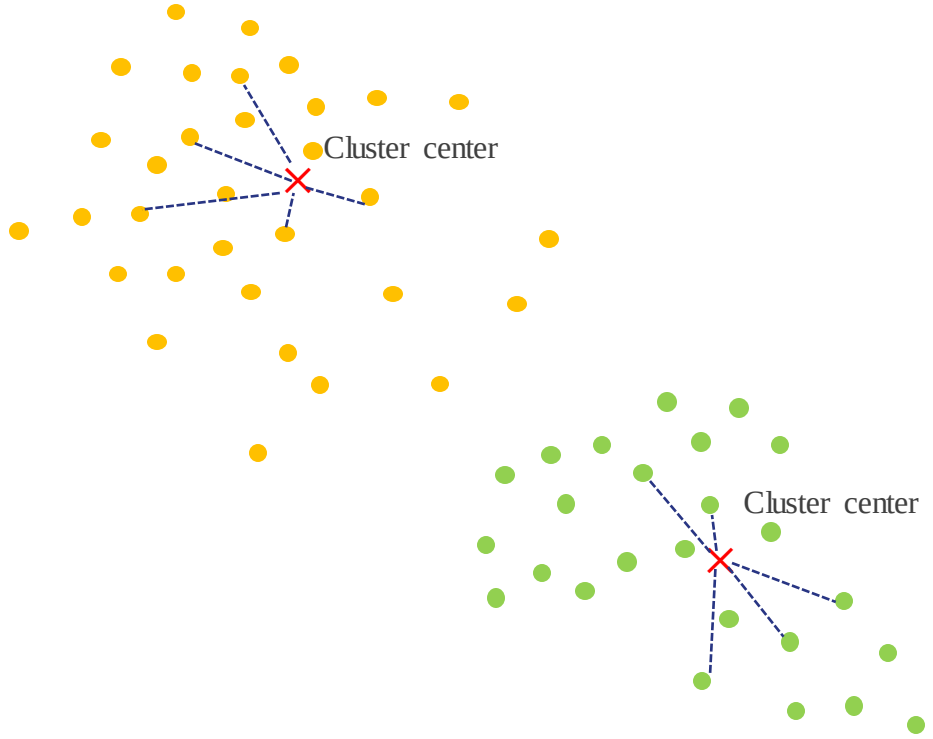
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Clustering Evaluation

- How to measure the clustering quality?

Elbow Technique





DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Clustering Evaluation

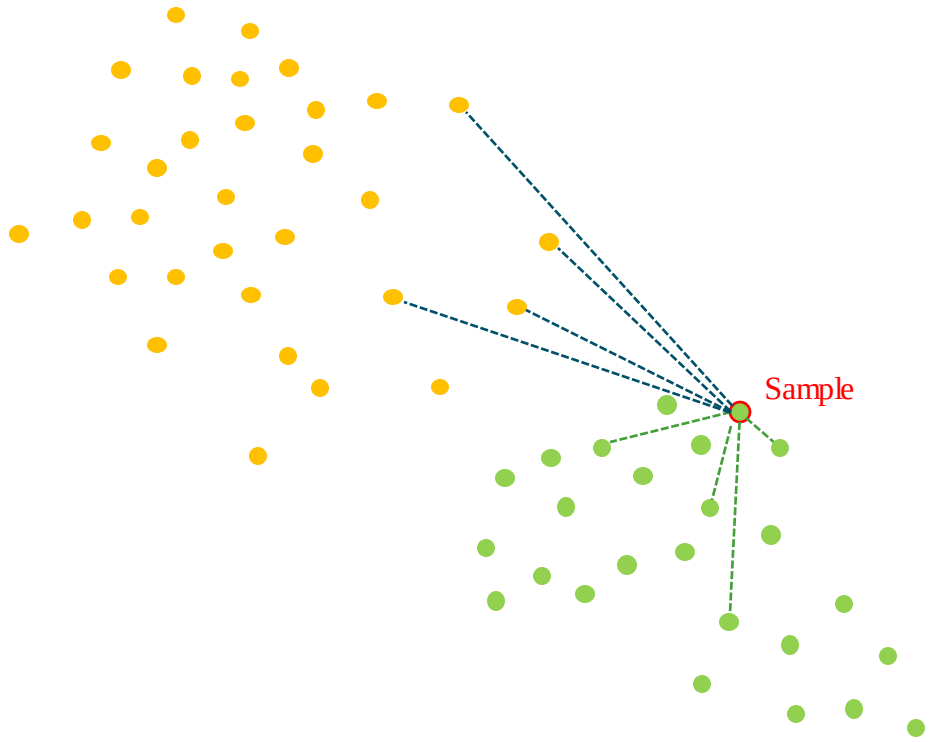
- How to measure the clustering quality?

Silhouette Coefficient:

$$s = \frac{b - a}{\max(a, b)}$$

a: The mean distance between a sample and all other points in the same cluster

b: The mean distance between a sample and all other points in the next nearest cluster



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Clustering Evaluation

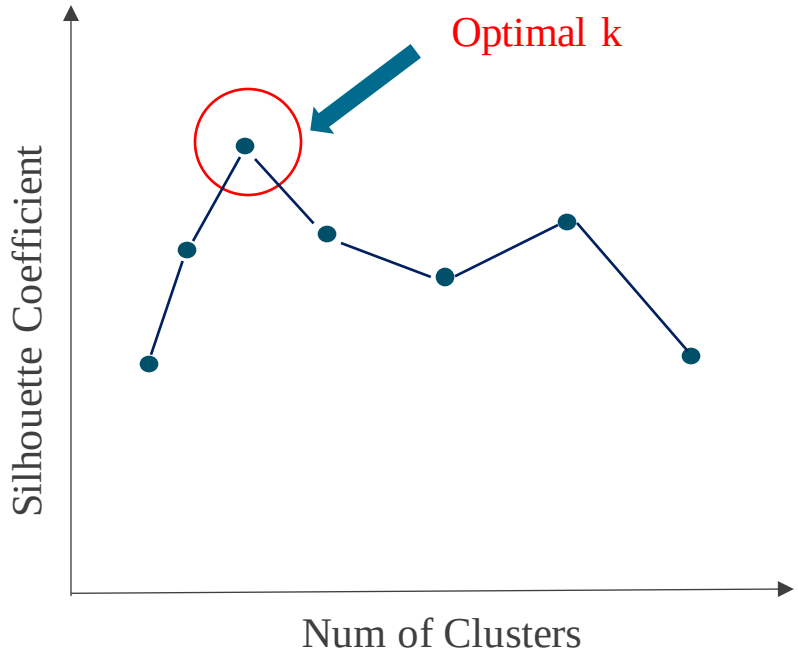
- How to measure the clustering quality?

Silhouette Coefficient:

$$s = \frac{b - a}{\max(a, b)}$$

a: The mean distance between a sample and all other points in the same cluster

b: The mean distance between a sample and all other points in the next nearest cluster



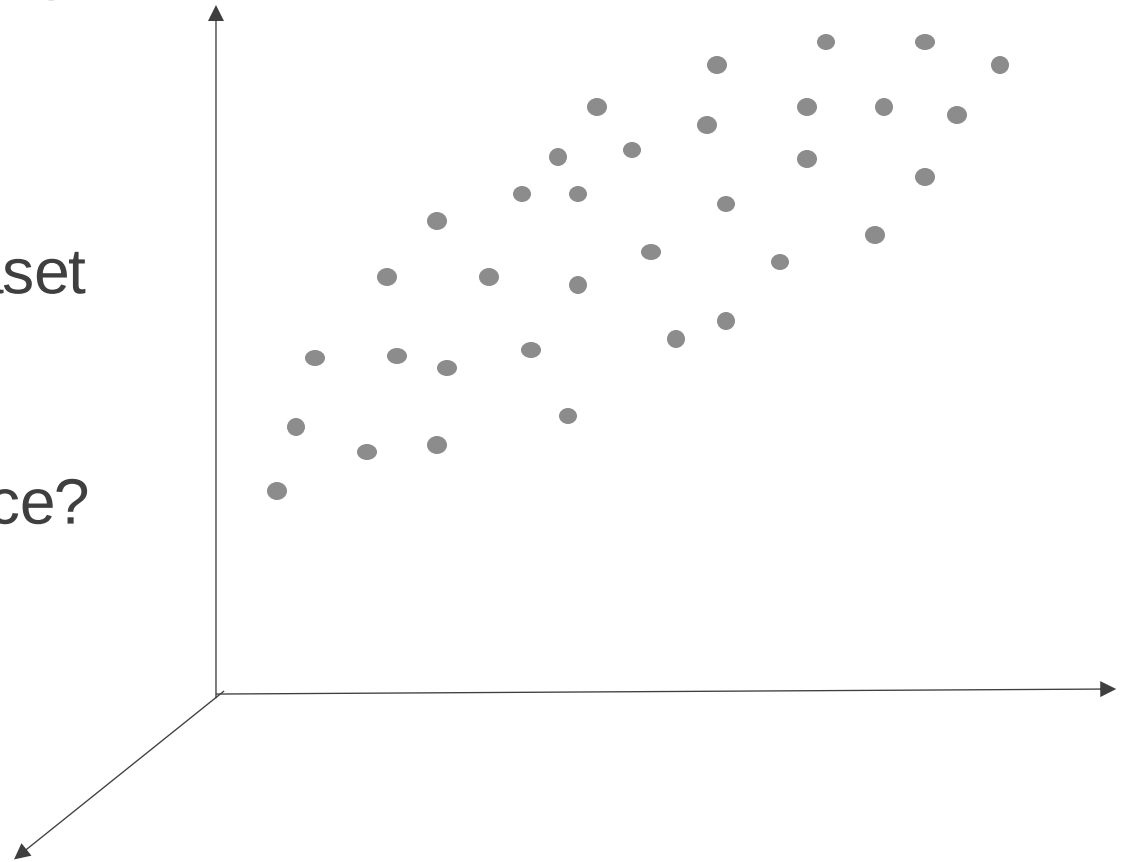
DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Dimension Reduction

- How to reduce dimensionality?
- Principal Component Analysis (PCA) most popular method

Think about a high dimensional dataset

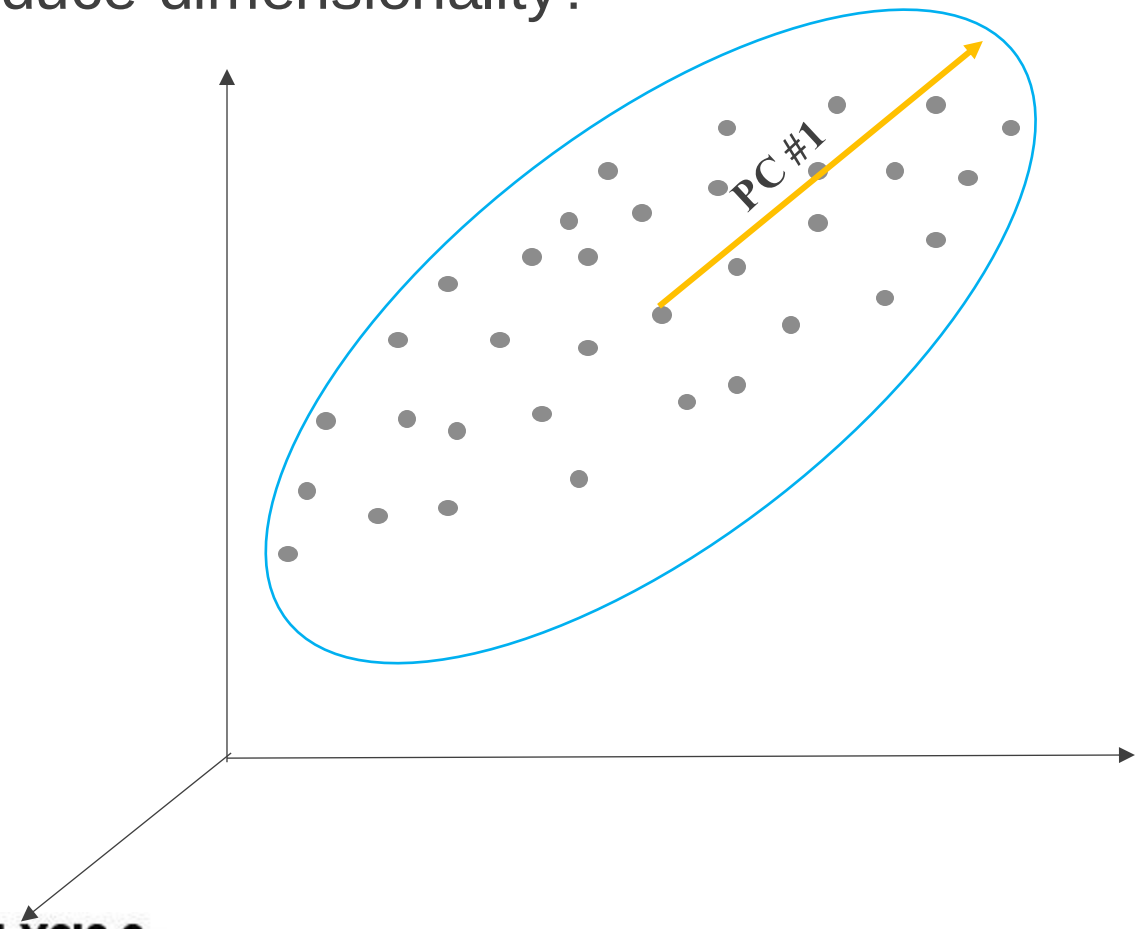
Which direction has the most variance?



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Dimension Reduction

- How to reduce dimensionality?

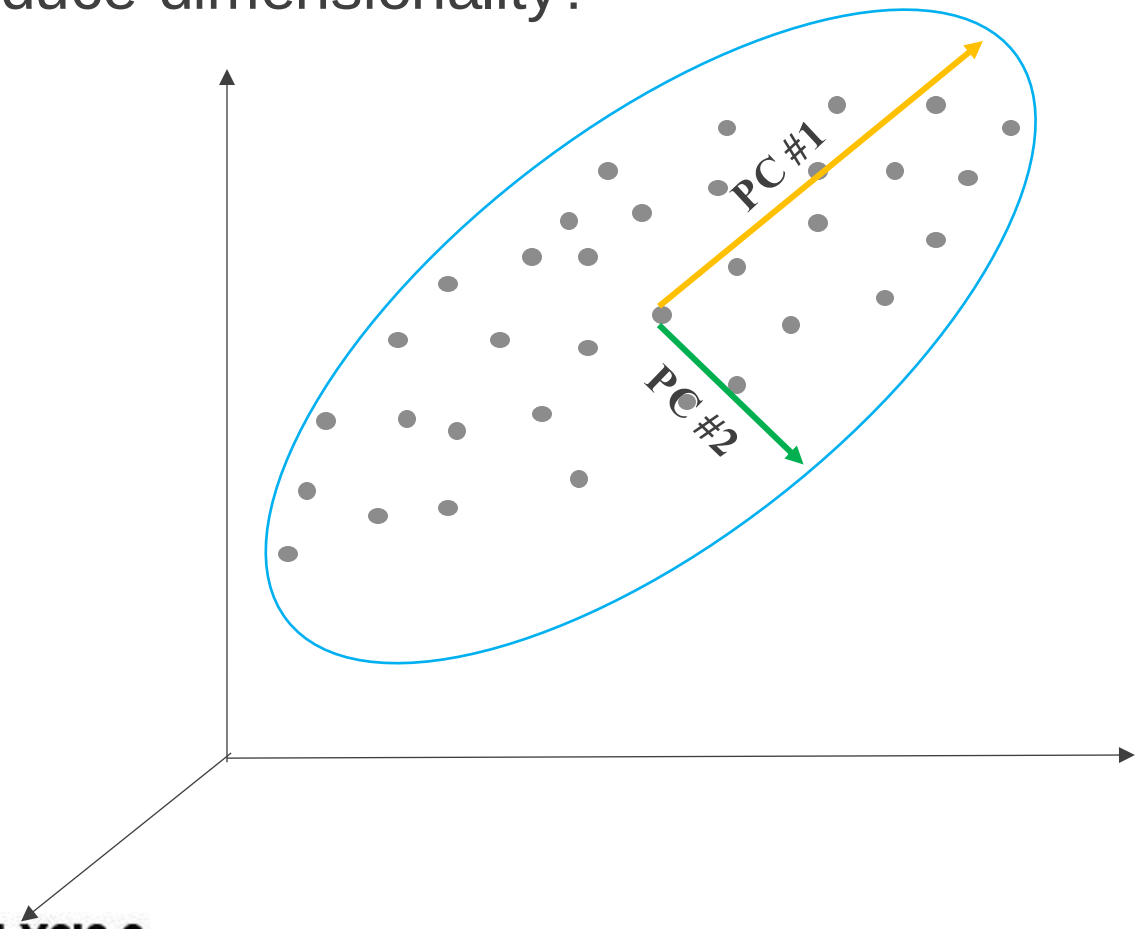


$$PC\#1 = 0.05X_1 - 0.84X_2 + \ldots + 0.11X_{200}$$

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Dimension Reduction

- How to reduce dimensionality?

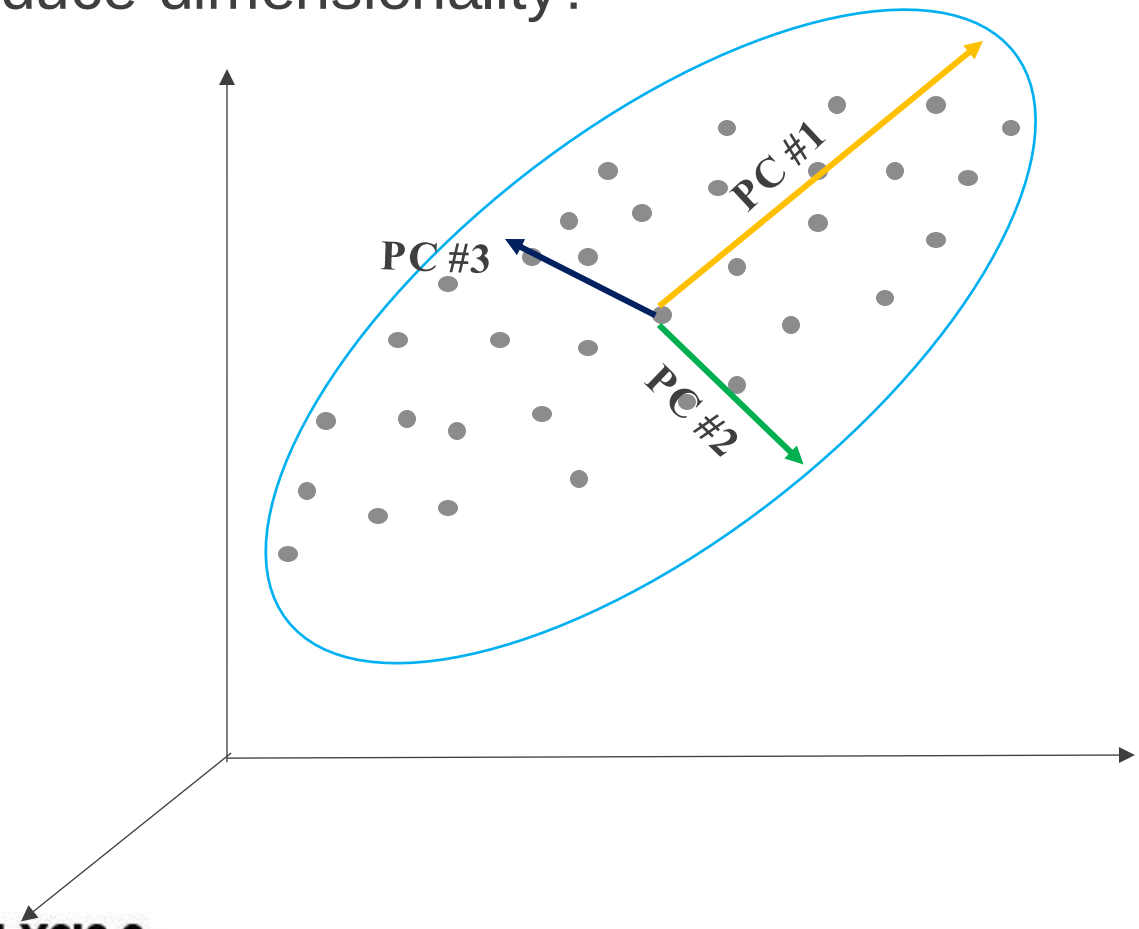


$$PC\#1 = 0.05X_1 - 0.84X_2 + \ldots + 0.11X_{200}$$
$$PC\#2 = 0.86X_1 + 0.05X_2 + \ldots - 0.47X_{200}$$

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Dimension Reduction

- How to reduce dimensionality?



$PC\#1 = 0.05X_1 - 0.84X_2 + \ldots + 0.11X_{200}$

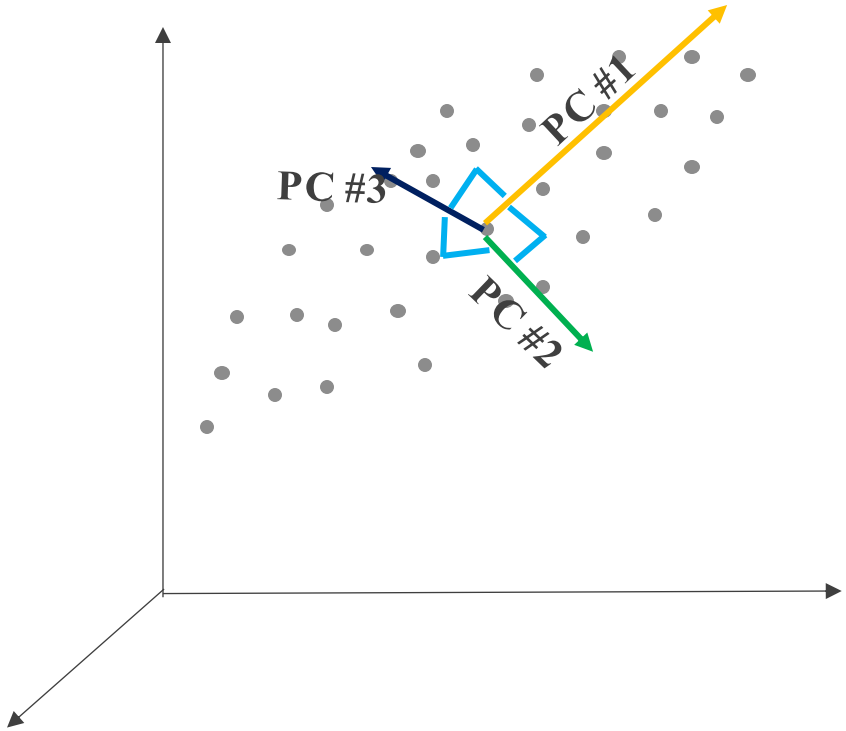
$PC\#2 = 0.36X_1 + 0.05X_2 + \ldots - 0.47X_{200}$

$PC\#3 = \ldots..$

DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Unsupervised Learning – Dimension Reduction

- Principal components are:
  - Orthogonal (or "perpendicular") to one another
  - Linear combinations of predictor variables
- # of Principal components = # of variables
  - Keep principal components that explain most of the variation -> reduce dimensionality



DATA ANALYSIS & SOCIAL INQUIRY LAB
GRINNELL COLLEGE

Q & A