

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



Outline for April 15

- Recap Handout 16
- Other NN architectures
- What is unsupervised learning?
- First unsupervised learning algorithm:
K-means clustering
 - Lab 7 due TODAY (last chance for late day)
 - Office hours TODAY 12:30-2pm
 - Proposal due FRIDAY

Outline for April 15

- **Recap Handout 16**
- Other NN architectures
- What is unsupervised learning?
- First unsupervised learning algorithm:
K-means clustering

Handout 16, Q1

- (a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)
- (b) First layer params
- (c) Second layer params
- (d) Third layer params
- (e) Total # params
- (f) Compare to FC with 312,860?

Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

CONV, FC

(b) First layer params

(c) Second layer params

(d) Third layer params

(e) Total # params

(f) Compare to FC with 312,860?

Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

CONV, FC

(b) First layer params $5*5*3*20 + 20 = 1520$

(c) Second layer params $3*3*20*10 + 10 = 1810$

(d) Third layer params $8*8*10*10 + 10 = 6410$

(e) Total # params 9740

(f) Compare to FC with 312,860?

Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

CONV, FC

(b) First layer params $5*5*3*20 + 20 = 1520$

(c) Second layer params $3*3*20*10 + 10 = 1810$

(d) Third layer params $8*8*10*10 + 10 = 6410$

(e) Total # params 9740

(f) Compare to FC with 312,860?

Much better! (over order of magnitude better)

Handout 16, Q2

(a) $W=10$, $F=7$, $P=3$, $S=3$

(b) Draw padding

(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding

(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding



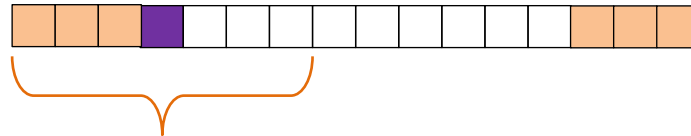
(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding



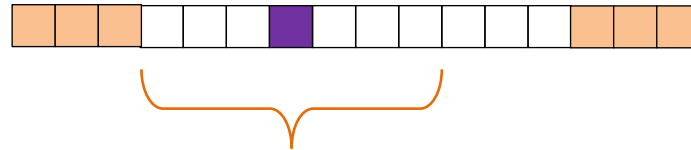
(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding



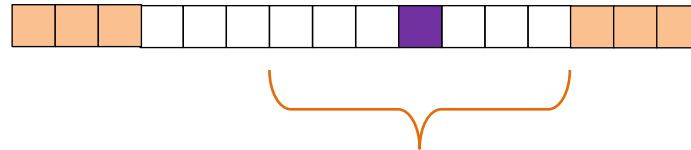
(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding



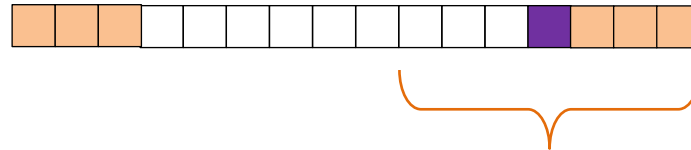
(c) Shade units where cross-correlation is performed

Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

$$(10-7+6)/3 + 1 = 4 \quad (\text{output size})$$

(b) Draw padding



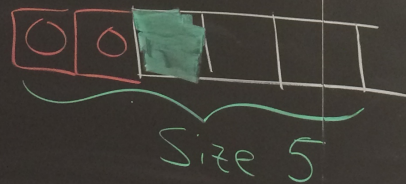
(c) Shade units where cross-correlation is performed

Handout 16, Q3

output
size
↓

$$32 = 1 + \frac{32 - 5 + 2P}{1}$$

$$P = 2$$



Q3:

$$W = 32$$

$$F = 5$$

$$S = 1$$














$$P = ?$$

Outline for April 15

- Recap Handout 16
- **Other NN architectures**
- What is unsupervised learning?
- First unsupervised learning algorithm:
K-means clustering

A mostly complete chart of Neural Networks

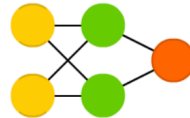
©2016 Fjodor van Veen - asimovinstitute.org

-  Backfed Input Cell
-  Input Cell
-  Noisy Input Cell
-  Hidden Cell
-  Probablistic Hidden Cell
-  Spiking Hidden Cell
-  Output Cell
-  Match Input Output Cell
-  Recurrent Cell
-  Memory Cell
-  Different Memory Cell
-  Kernel
-  Convolution or Pool

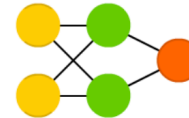
Perceptron (P)



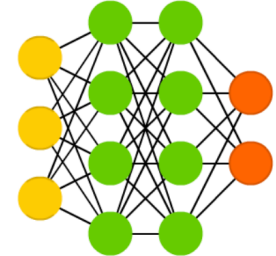
Feed Forward (FF)



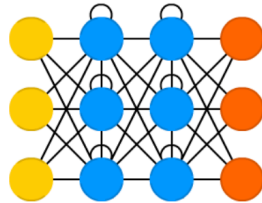
Radial Basis Network (RBF)



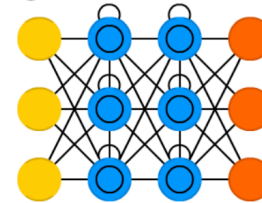
Deep Feed Forward (DFF)



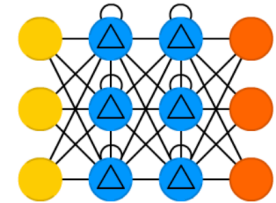
Recurrent Neural Network (RNN)



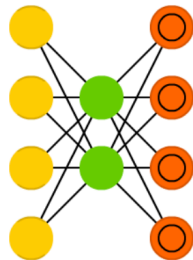
Long / Short Term Memory (LSTM)



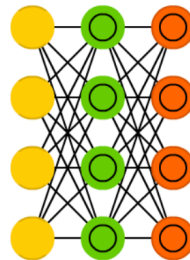
Gated Recurrent Unit (GRU)



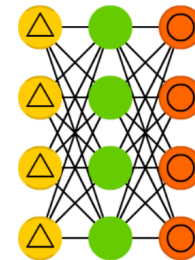
Auto Encoder (AE)



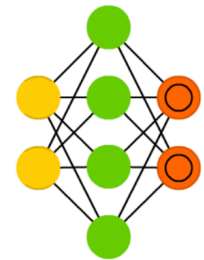
Variational AE (VAE)

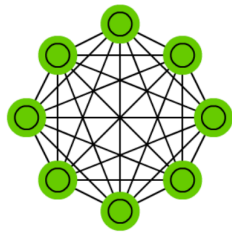


Denoising AE (DAE)

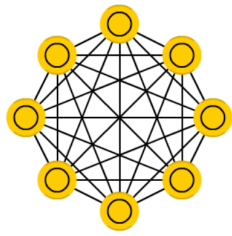


Sparse AE (SAE)

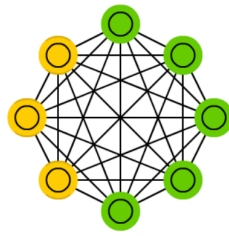




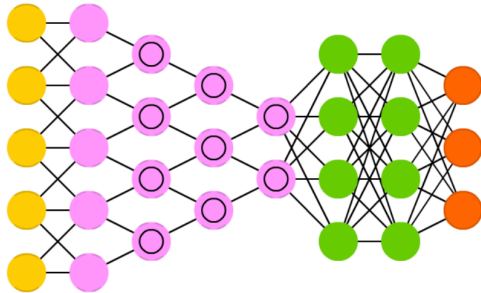
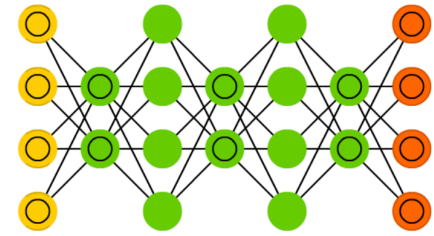
Deep Convolutional Network (DCN)



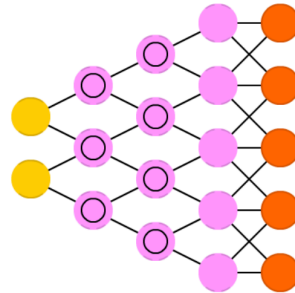
Deconvolutional Network (DN)



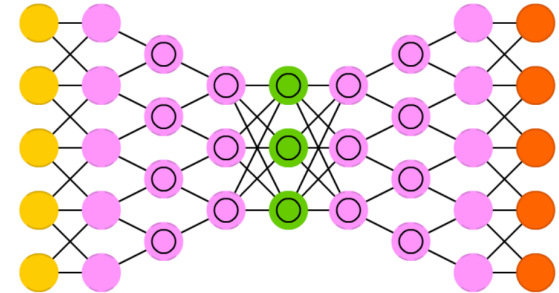
Deep Convolutional Inverse Graphics Network (DCIGN)



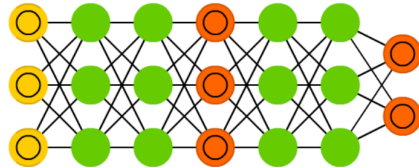
Generative Adversarial Network (GAN)



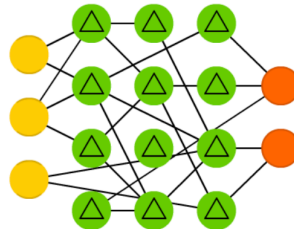
Liquid State Machine (LSM)



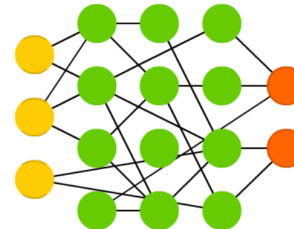
Extreme Learning Machine (ELM)



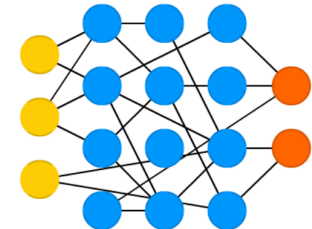
Deep Residual Network (DRN)



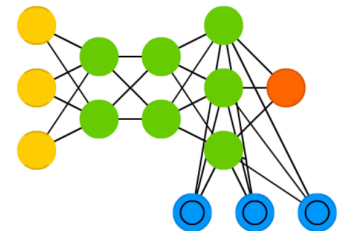
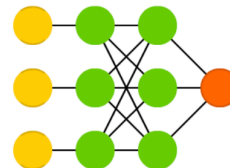
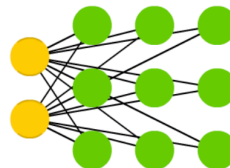
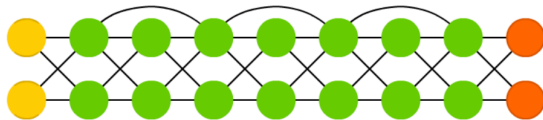
Kohonen Network (KN)



Support Vector Machine (SVM)



Neural Turing Machine (NTM)



Generative Adversarial Networks

- Idea: have two “adversarial” NNs

Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data

Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data
- The other network (“art detective”) is trying to determine whether or not produced works are “fakes”

Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data
- The other network (“art detective”) is trying to determine whether or not produced works are “fakes”
- As training proceeds, the forger becomes better at producing works in a given style

Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data
- The other network (“art detective”) is trying to determine whether or not produced works are “fakes”
- As training proceeds, the forger becomes better at producing works in a given style
- But the art detective becomes better at discriminating between originals and fakes

GAN progress over time



Ian Goodfellow

@goodfellow_ian

Follow



4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661
arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536
arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948



4:40 PM - 14 Jan 2019

GAN painting

Sold for almost half
a million dollars



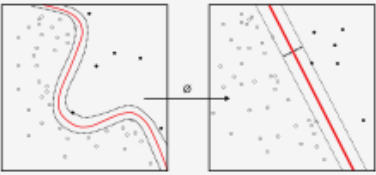
Outline for April 15

- Recap Handout 16
- Other NN architectures
- **What is unsupervised learning?**
- First unsupervised learning algorithm:
K-means clustering

Supervised Learning:

makes use of examples where we know the underlying “truth” (label/output)

Machine learning and data mining



Problems [show]

Supervised learning [hide]
(classification • regression)

Decision trees • Ensembles (Bagging, Boosting, Random forest) • *k*-NN • Linear regression • Naive Bayes • Neural networks • Logistic regression • Perceptron • Relevance vector machine (RVM) • Support vector machine (SVM)

Clustering [hide]
BIRCH • **Hierarchical** • *k*-means • Expectation-maximization (EM) • DBSCAN • OPTICS • Mean-shift

Dimensionality reduction [hide]
Factor analysis • CCA • ICA • LDA • NMF • PCA • t-SNE

Structured prediction [hide]
Graphical models (Bayes net, CRF, HMM)


Anomaly detection [hide]
k-NN • Local outlier factor

Neural nets [hide]
Autoencoder • Deep learning • Multilayer perceptron • RNN • Restricted Boltzmann machine • SOM • Convolutional neural network

Reinforcement Learning [hide]
Q-Learning • SARSA • Temporal Difference (TD)

Theory [show]

Machine learning venues [show]

 **Machine learning portal**

V • T • E

Unsupervised Learning:

Learn underlying structure or features without labeled training data

Unsupervised learning: 3 main areas

- 1) Clustering: group data points into clusters based on features only
- 2) Dimensionality reduction: remove feature correlation, compress data, visualize data
- 3) Structured prediction: model latent variables (example: Hidden Markov Models)

Unsupervised learning examples from biology: clustering

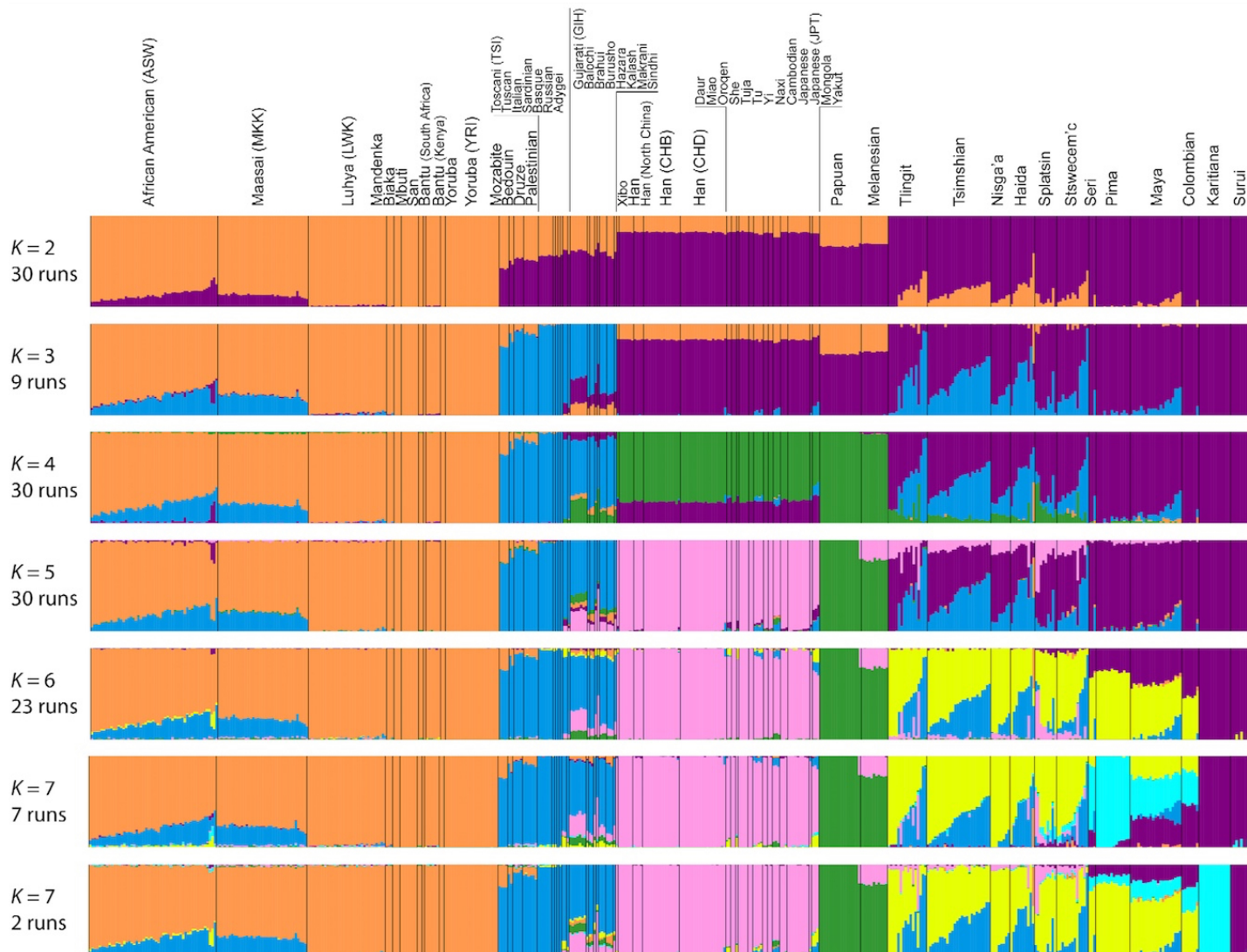
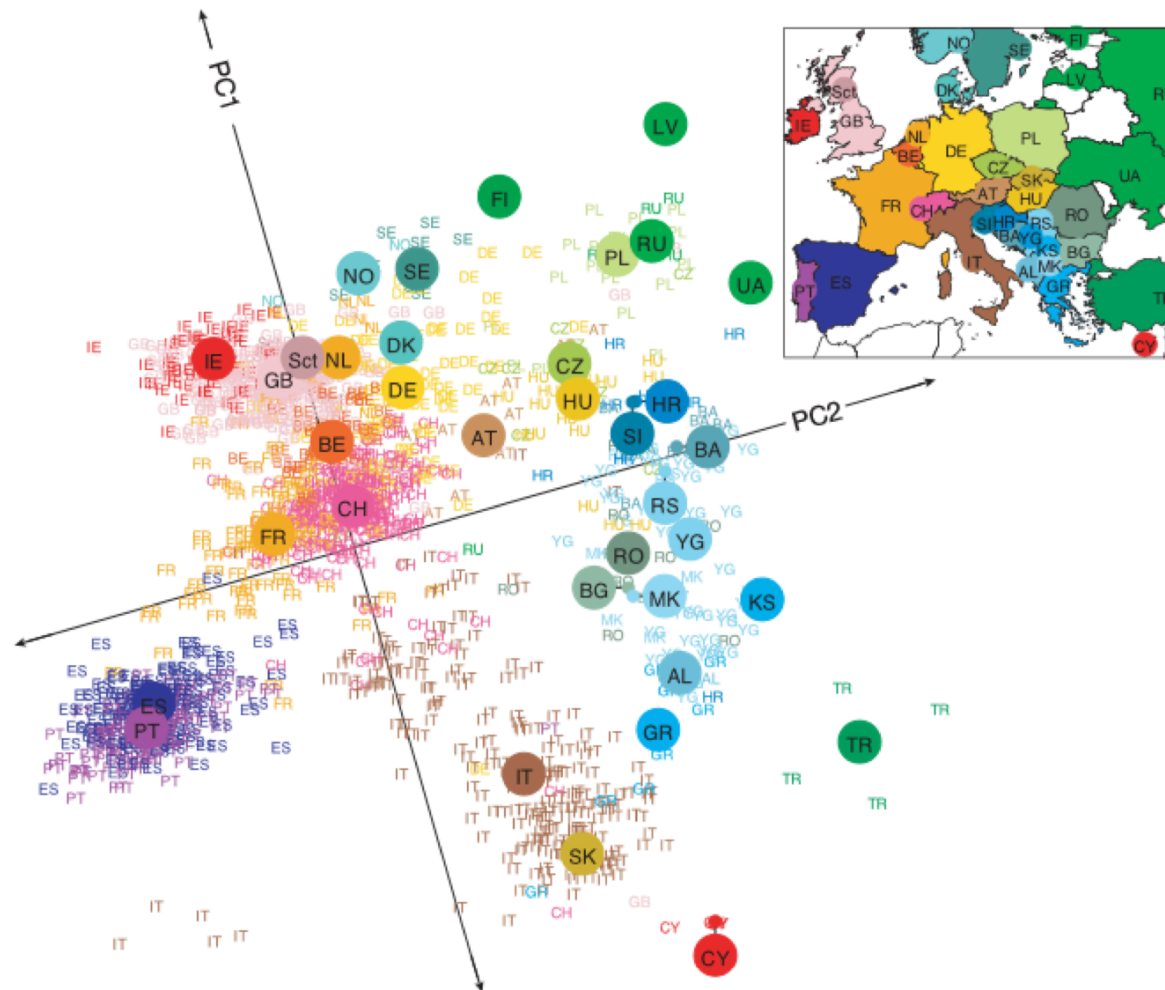
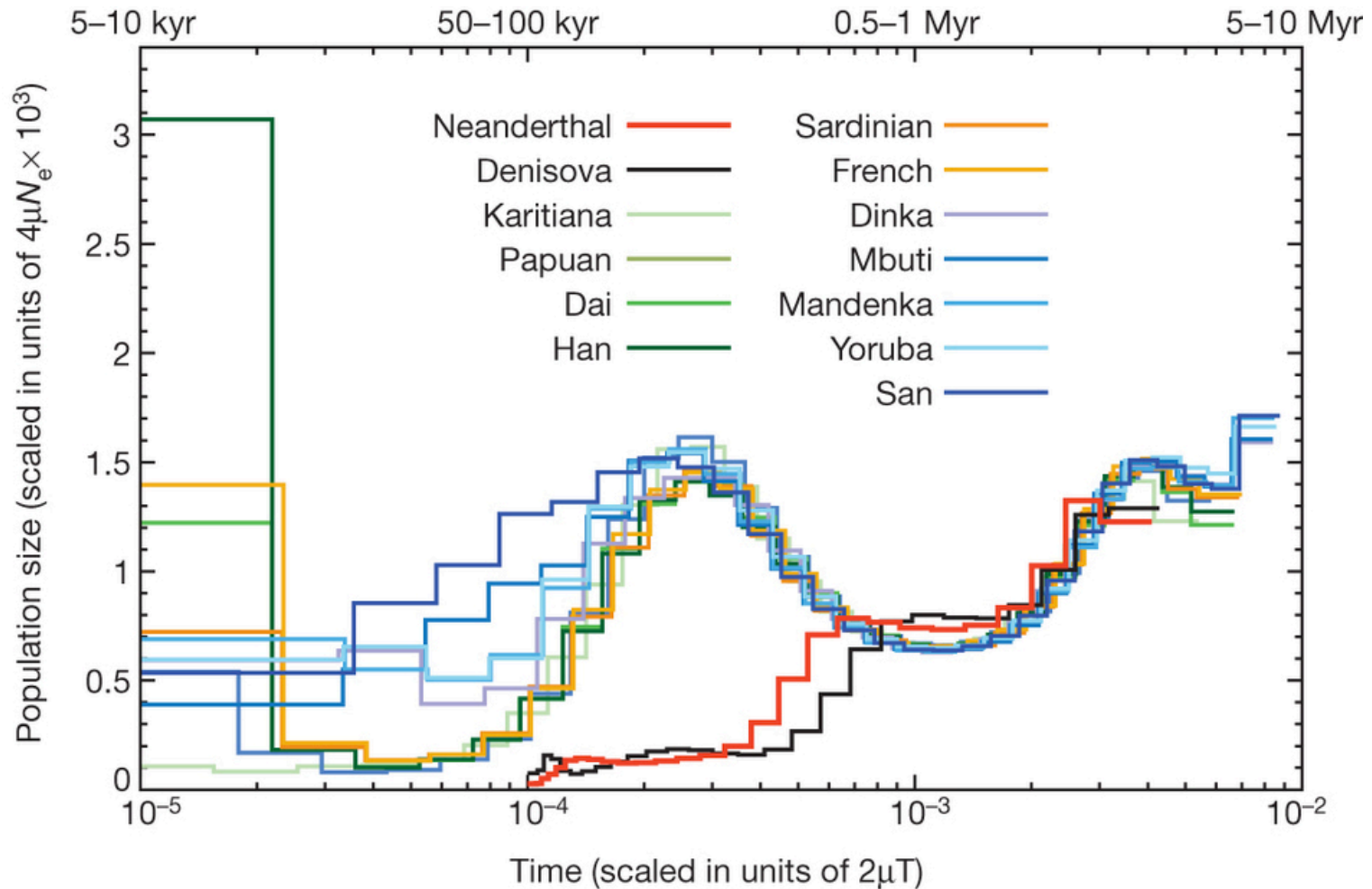


Figure: German Dziebel

Unsupervised learning examples from biology: structured prediction



Unsupervised learning examples from biology: structured prediction



Outline for April 15

- Recap Handout 16
- Other NN architectures
- What is unsupervised learning?
- First unsupervised learning algorithm:
K-means clustering

K-means

Goal: find K clusters

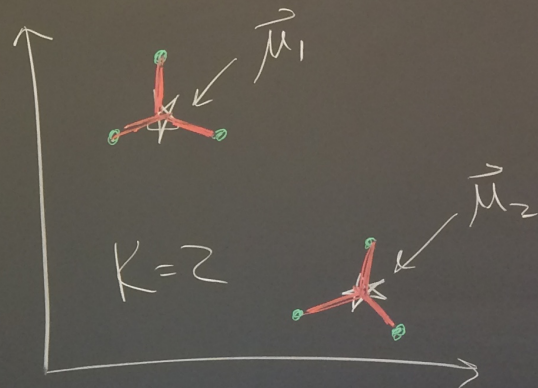
$\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K\} = \mathcal{C}$, with means $\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_K$

s.t.

minimizes

$$J(\mathcal{C}) = \sum_{k=1}^K \sum_{\vec{x}_i \in \mathcal{C}_k} \|\vec{x}_i - \vec{\mu}_k\|^2$$

Within cluster sum of squares



red distances =
within cluster
sum of squares

Algo

Iterate
[E]

Algorithm

Initialization

Choose $\vec{\mu}_1^{(1)}, \vec{\mu}_2^{(1)}, \dots, \vec{\mu}_k^{(1)}$
(usually points from training data)

Iterate

E-step

Assign each point \vec{x} a cluster k based on the min distance to each cluster mean $\vec{\mu}_k^{(t)}$
 $\Rightarrow \vec{x} \in C_k^{(t)}$

expected

M-step

Update cluster means

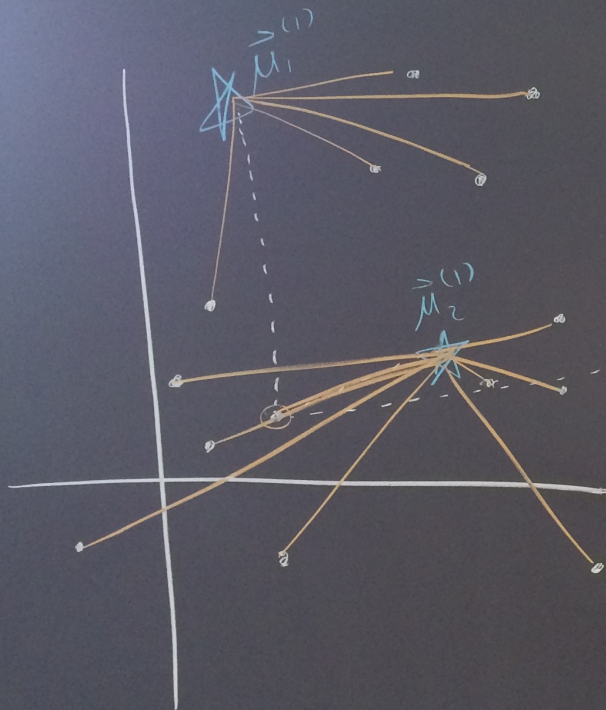
$$\vec{\mu}_k^{(t+1)} = \frac{1}{|C_k^{(t)}|} \sum_{\vec{x}_i \in C_k^{(t)}} \vec{x}_i \quad \left(\begin{array}{l} \text{avg} \\ \text{of points} \\ \text{assigned} \\ \text{to each} \\ \text{cluster} \end{array} \right)$$

Stop

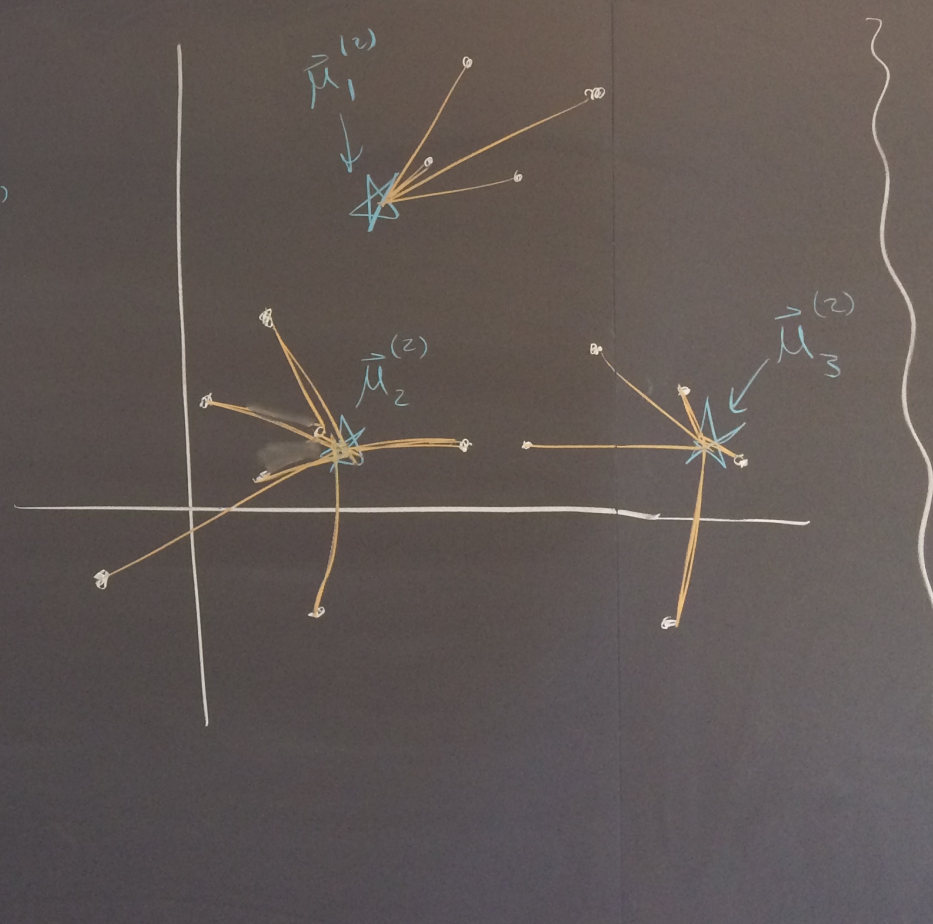
- * no cluster membership changes (means don't change)
- * max iter exceeded.
- * see config you saw before (cycle)

Example with $p=2$, $n=16$

Iteration 1



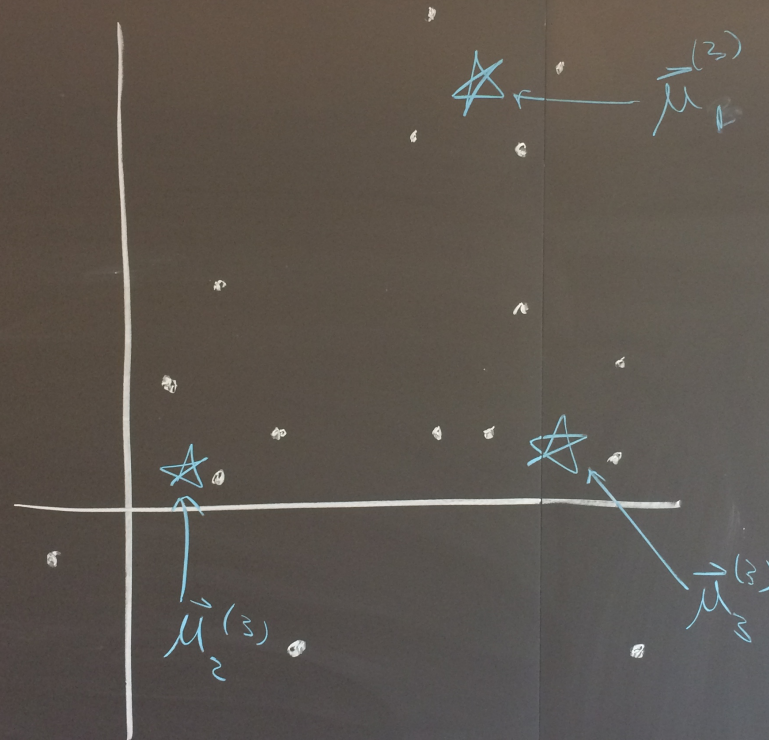
Iteration 2



Note: normally we would choose initial means from among training data point, but here we did not

Example with $p=2$, $n=16$

Iteration 3



① runtime? K, n, p

② how to choose K ?

NEXT TIME!