CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2020

HAVERFORD
COLLEGE

Admin

- Office hours **today 11-12pm**

- **Project**
  - Proposals were due last night – email me ASAP if you have not yet submitted one and would like to do the project
  - I will grade *one* of project or exam, so if you do both email me by the end of the day on Dec 18 which one to grade

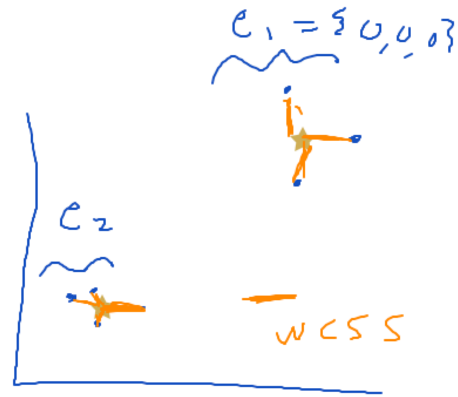
Outline for December 4

- Continue K-means clustering
- Handout 18
- Gaussian Mixture Models (GMM)
- Next week:
  - Principal Component Analysis (PCA)
  - Review session
  - Presentations
  - Capstone activity

Outline for December 4

- Continue K-means clustering
- Handout 18
- Gaussian Mixture Models (GMM)
- Next week:
  - Principal Component Analysis (PCA)
  - Review session
  - Presentations
  - Capstone activity

Clustering Goals



- Learn about underlying structure in the data
- Cluster new data (testing)

- Goal: minimize "within cluster sum of squares" (WCSS)

find $\{C_1, C_2, \dots C_k\} = C$ s.t.

WCSS (cost) $\rightarrow$

$$J(C) = \underbrace{\sum_{k=1}^{K}}_{\text{all clusters}} \sum_{i \in C_k} \| \vec{x}_i - \vec{\mu}_k \|^2$$

train data

cluster mean

truly minimizing WCSS

$\Rightarrow$ NP-hard

K-means algorithm

n, p

- 0) Initialization: choose means (centers) of K clusters ← typically from training data

$\vec{\mu}_1^{(1)}, \vec{\mu}_2^{(1)}, \ldots, \vec{\mu}_K^{(1)}$

- 1) **E-step (assignment):** for each training example, find closest mean

$\vec{x}$ → label k, $\vec{x} \in C_k^{(t)}$

repeat until convergence

- 2) **M-step (update):** compute new means

$$\mu_k^{(t)} = \frac{1}{|C_k^{(t)}|} \sum_{\vec{x}_i \in C_k^{(t)}} \vec{x}_i$$

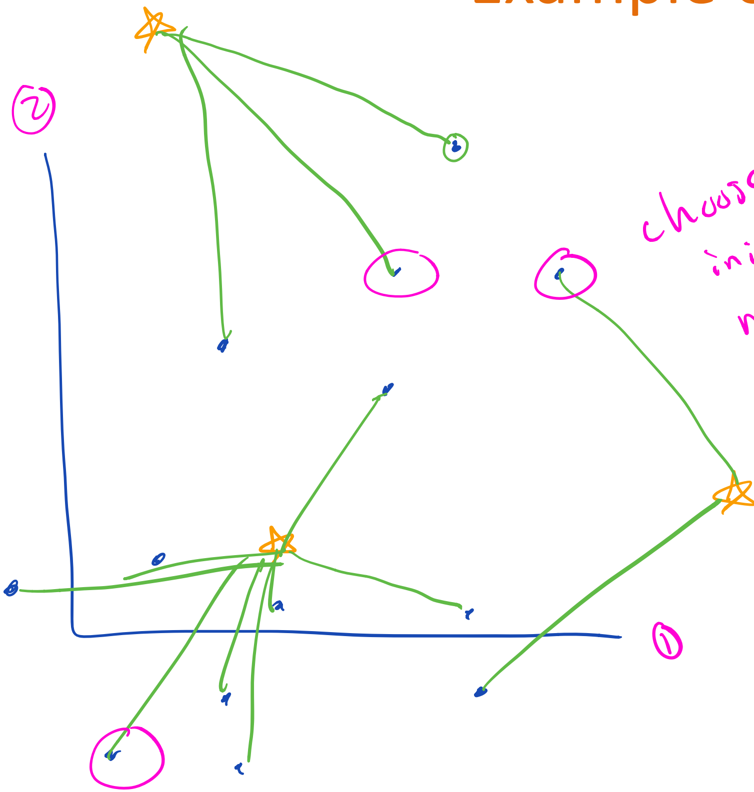
K-means algorithm: stopping criteria

- No cluster membership changes
- Max iterations exceeded
- Seen a configuration we’ve seen before (cycle)

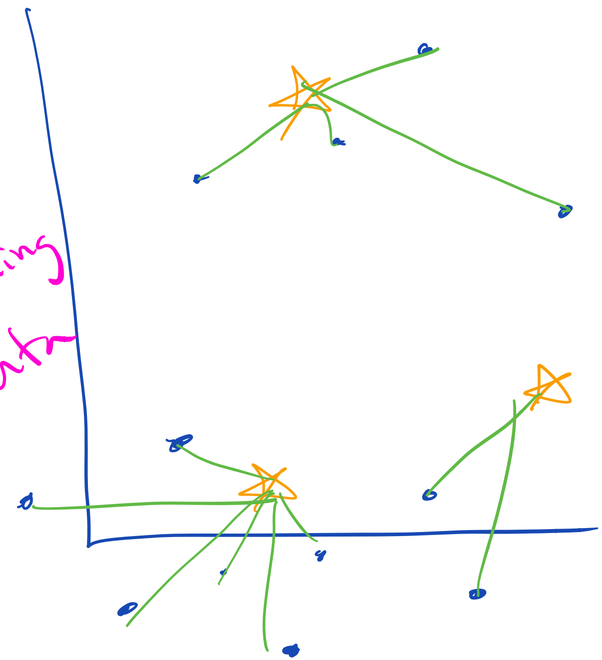
Example of K-means

3

$p = 2$



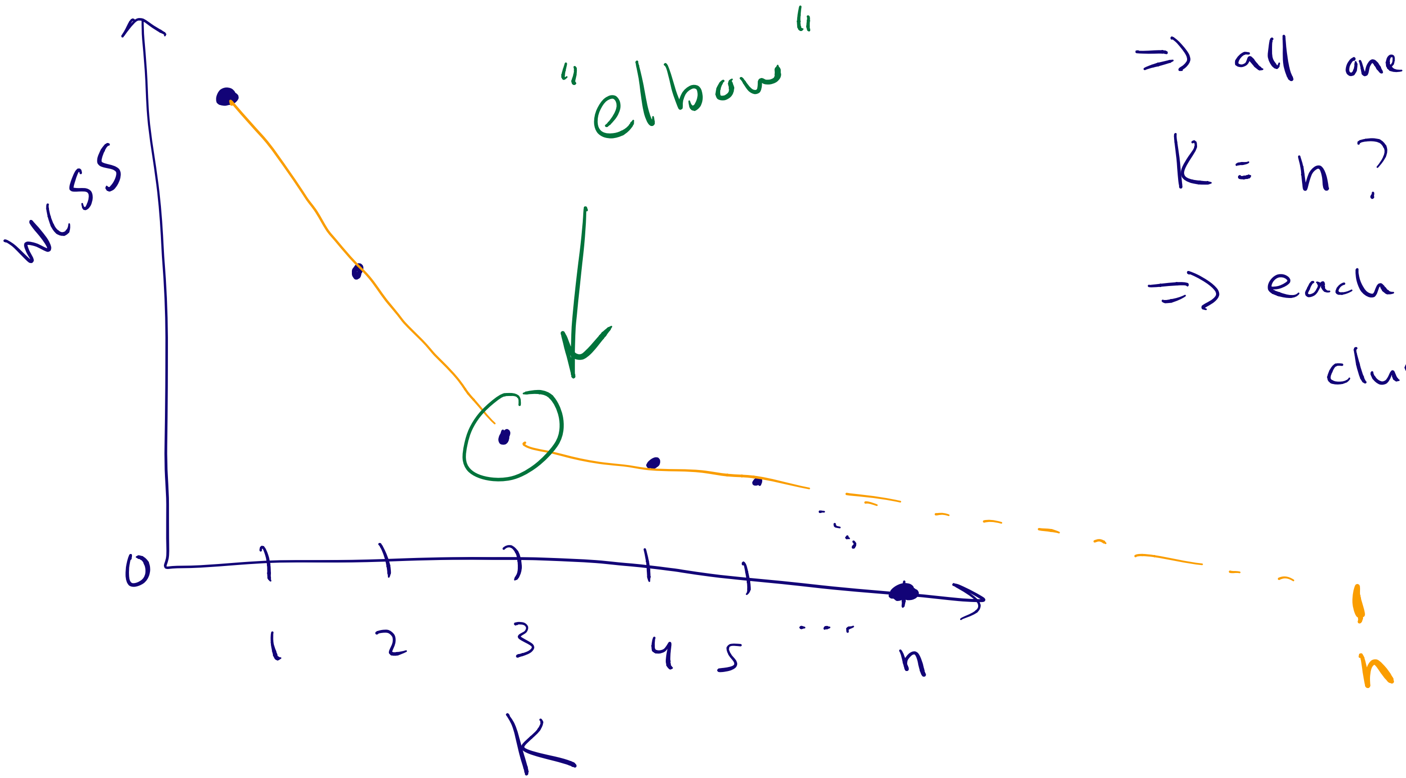
choose initial means from training data



WCSS —

↳ minimized over the iterations

How to choose K?



$\Rightarrow$ $\boxed{K = 3}$

K = 1 ?

$\Rightarrow$ all one cluster

K = n ?

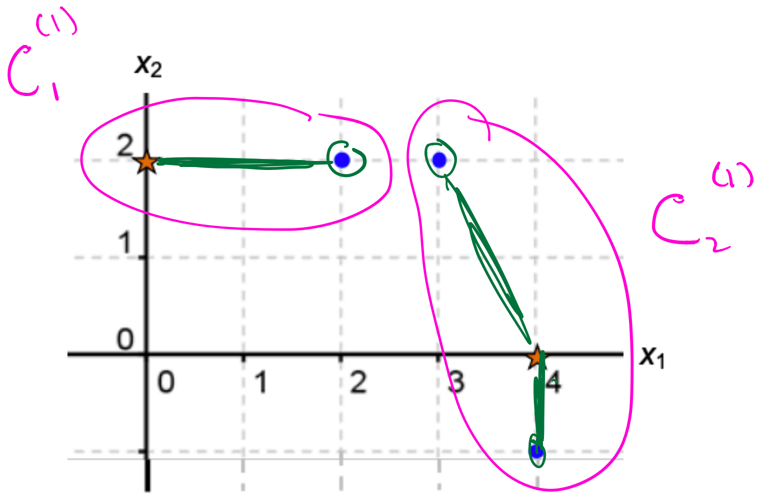
$\Rightarrow$ each own cluster

Outline for December 4

- Continue K-means clustering
- Handout 18
- Gaussian Mixture Models (GMM)
- Next week:
  - Principal Component Analysis (PCA)
  - Review session
  - Presentations
  - Capstone activity

1. Consider the data below with $n = 3$ and $p = 2$. The graph below shows these 3 points (circles), as well as the initial means (stars) for $K = 2$. Here $\vec{\mu}_1^{(1)} = [0, 2]$ and $\vec{\mu}_2^{(1)} = [4, 0]$.

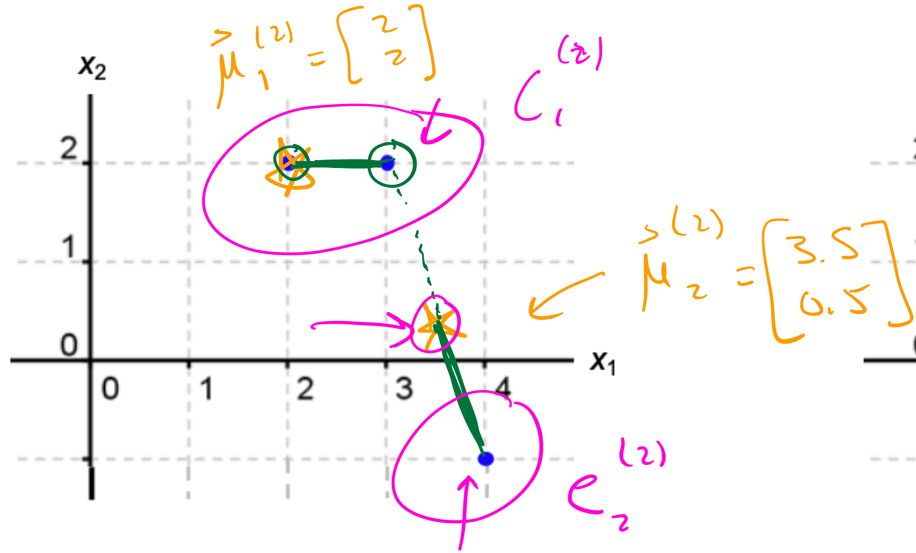
$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \\ 4 & -1 \end{bmatrix}$$

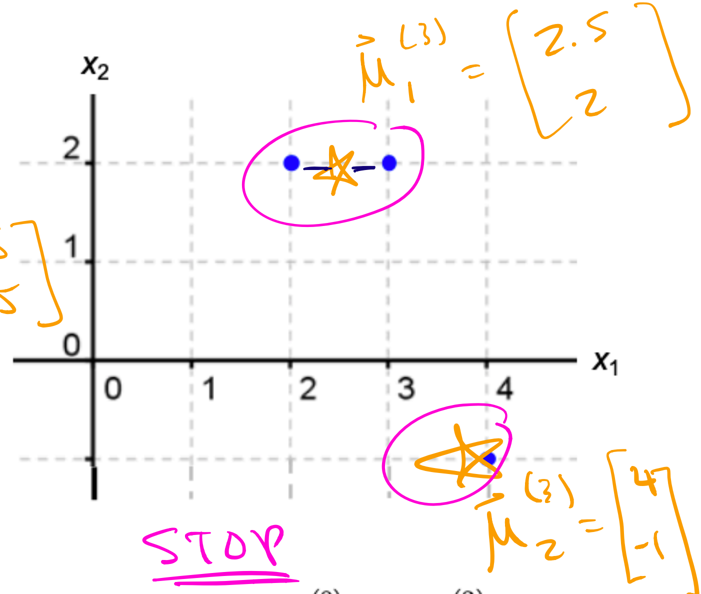


(a) On the graph above, show the cluster membership of each point, based on these initial means. What are $\mathcal{C}_1^{(1)}$ and $\mathcal{C}_2^{(1)}$?

$C_1^{(1)} = \{\vec{x}_2\}$ $\quad$ $C_2^{(1)} = \{\vec{x}_1, \vec{x}_3\}$

(b) Based on these cluster memberships, what are $\vec{\mu}_1^{(2)}$ and $\vec{\mu}_2^{(2)}$? Draw these two points as stars on the left plot below. This concludes the first iteration of the K-means algorithm.





(c) Based on the new means, draw the new cluster memberships and list $\mathcal{C}_1^{(2)}$ and $\mathcal{C}_2^{(2)}$. Finally, on the right plot above, draw the final means $\vec{\mu}_1^{(3)}$ and $\vec{\mu}_2^{(3)}$ and write out their values.

2. Does the "within cluster sum of squares" (WCSS) always decrease as K (number of clusters) increases?

yes (monotonic)

3. Compute the WCSS for the points above, using $K = 1$, $K = 2$, and $K = 3$.

$K=1$ $\vec{\mu} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

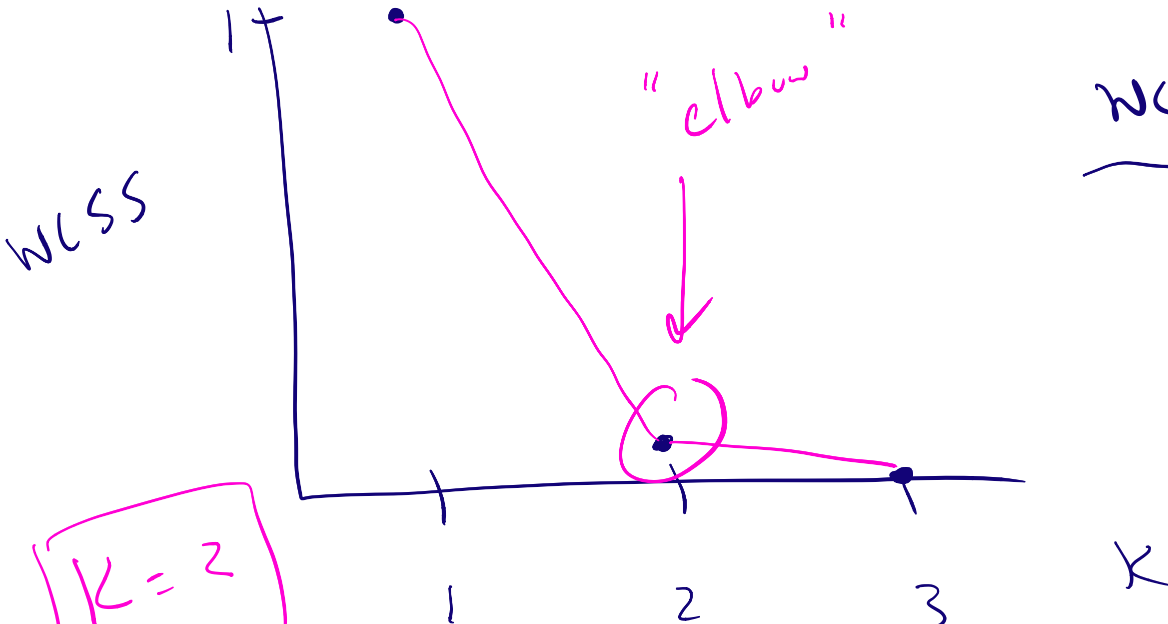
$WCSS(1) = 8$

$K=2$

$WCSS(2) = \frac{1}{2}$

$K=3$

$WCSS(3) = 0$



$\boxed{K=2}$

4. Finally, plot $K = 1, 2, 3$ on the x-axis and WCSS on the y-axis to create an "elbow" plot. What K would you choose in this case?

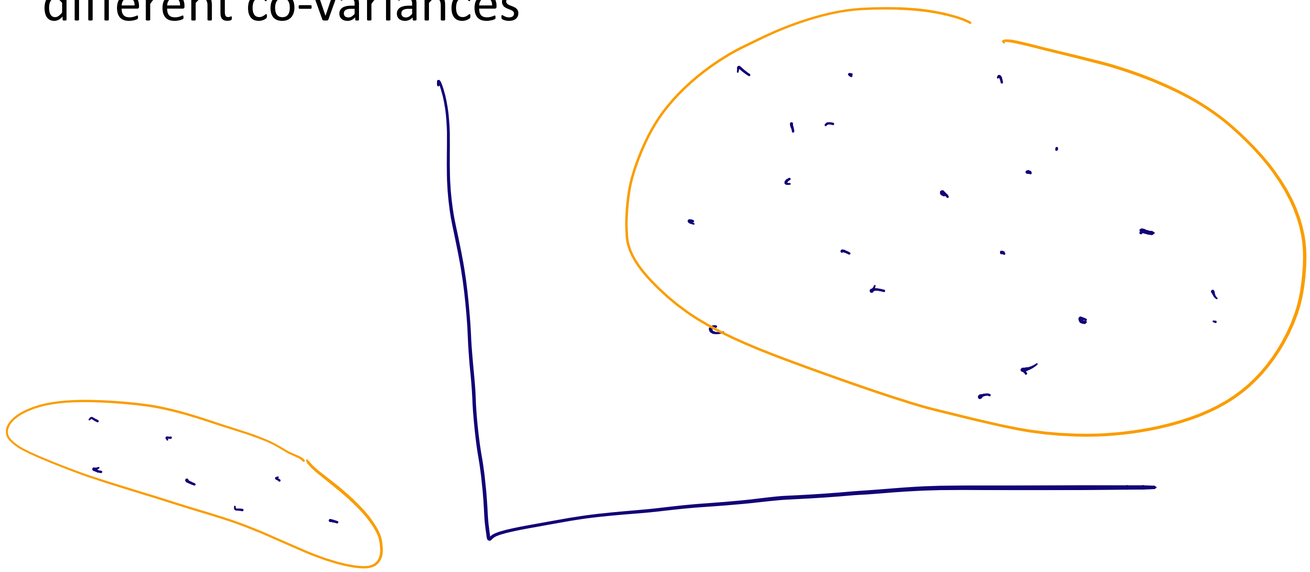
5. In terms of n, p, K, and T (max number of iterations), what is the runtime of K-means?

Problems with K-means

- Not generative (could not create a new point based on existing clusters)

- Does not explicitly account for different cluster sizes

- Implicitly assumes “spherical” clusters; does not account for different co-variances



Outline for December 4

- Continue K-means clustering
- Handout 18
- Gaussian Mixture Models (GMM)
- Next week:
  - Principal Component Analysis (PCA)
  - Review session
  - Presentations
  - Capstone activity

Discriminative vs. Generative

- Discriminative: finds a decision boundary
  - Logistic regression, K-means
- Generative: estimates probability distributions
  - Naïve Bayes, Gaussian Mixture Models

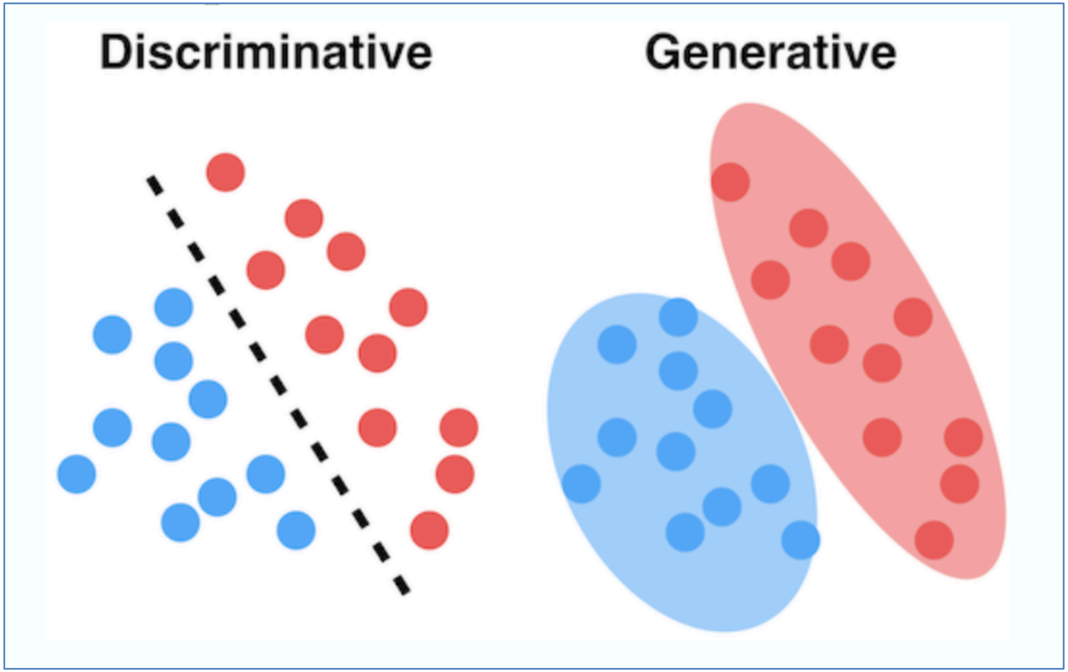
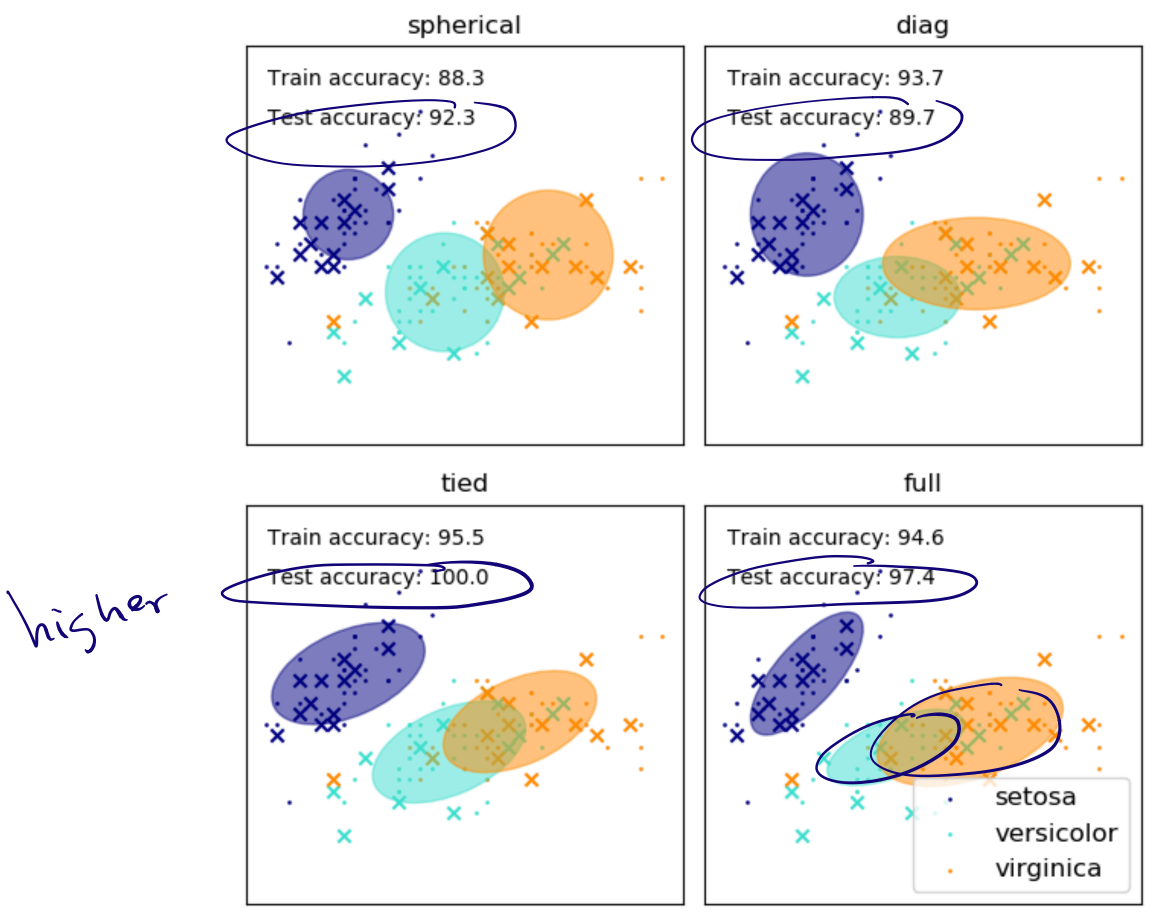


Figure: Ameet Soni

Example of GMMs with different covariance constraints on the Iris flower data



https://scikit-learn.org/stable/auto_examples/mixture/plot_gmm_covariances.html#sphx-glr-auto-examples-mixture-plot-gmm-covariances-py

Gaussian Mixture Model (GMM) ← K clusters

Likelihood

π_k = prob. of class k (cluster size)

hidden cluster

ex: $K=3$, $[\pi_1, \pi_2, \pi_3]$

$= [1/6, 1/2, 1/3]$

$$p(\vec{x}) = \sum_{k=1}^{K} p(\vec{x}, z=k)$$

$p(z=k)$ Gaussian

$$= \sum_{k=1}^{K} \underbrace{\pi_k \, p(\vec{x} \mid z=k)}_{p(A)\,p(B \mid A)}$$

model

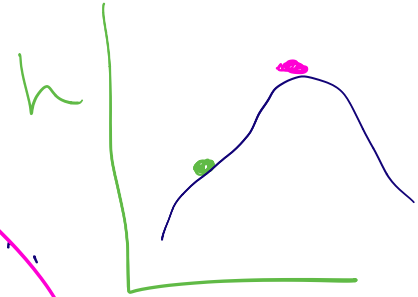
$$\underbrace{L(X)}_{\text{max}} = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k \, N(\vec{x}_i, \vec{\mu}_k, \sigma_k^2)$$

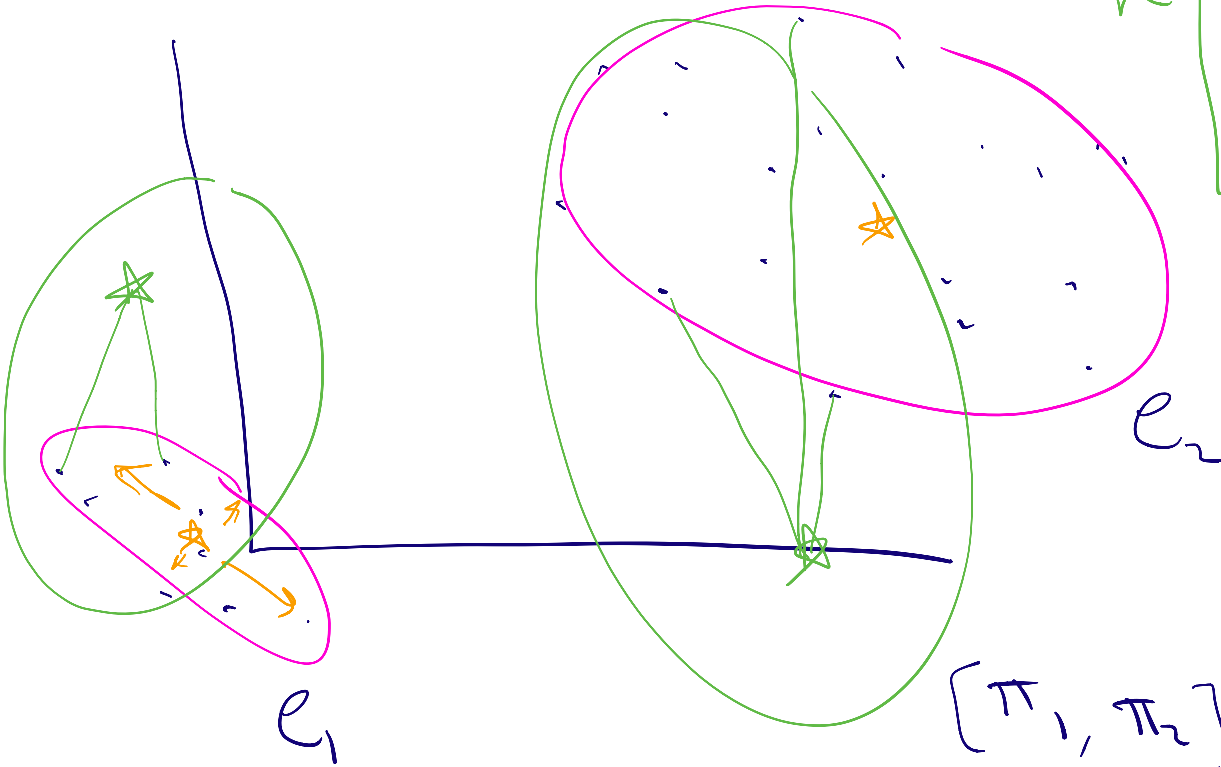
solve for

mean

variance

EM algorithm (high-level)





end result

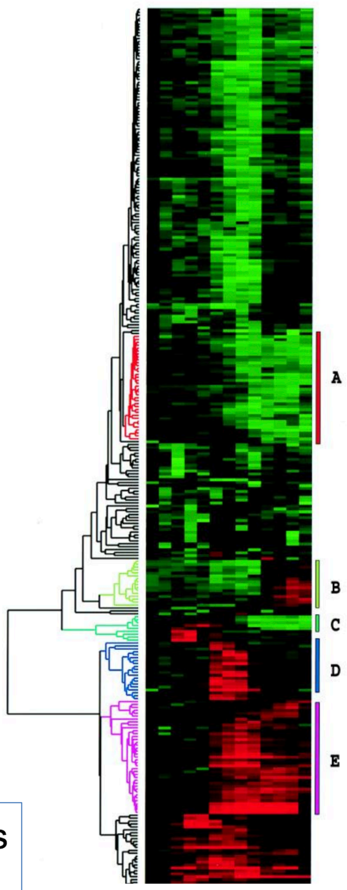
$$[\pi_1, \pi_2] = \underbrace{[0.2, 0.8]}_{\text{end result}}$$

EM algorithm (high-level)

Bonus: Hierarchical Clustering

Applications of clustering in ML

- Cluster genes with similar expression patterns



Cluster analysis and display of genome-wide expression patterns

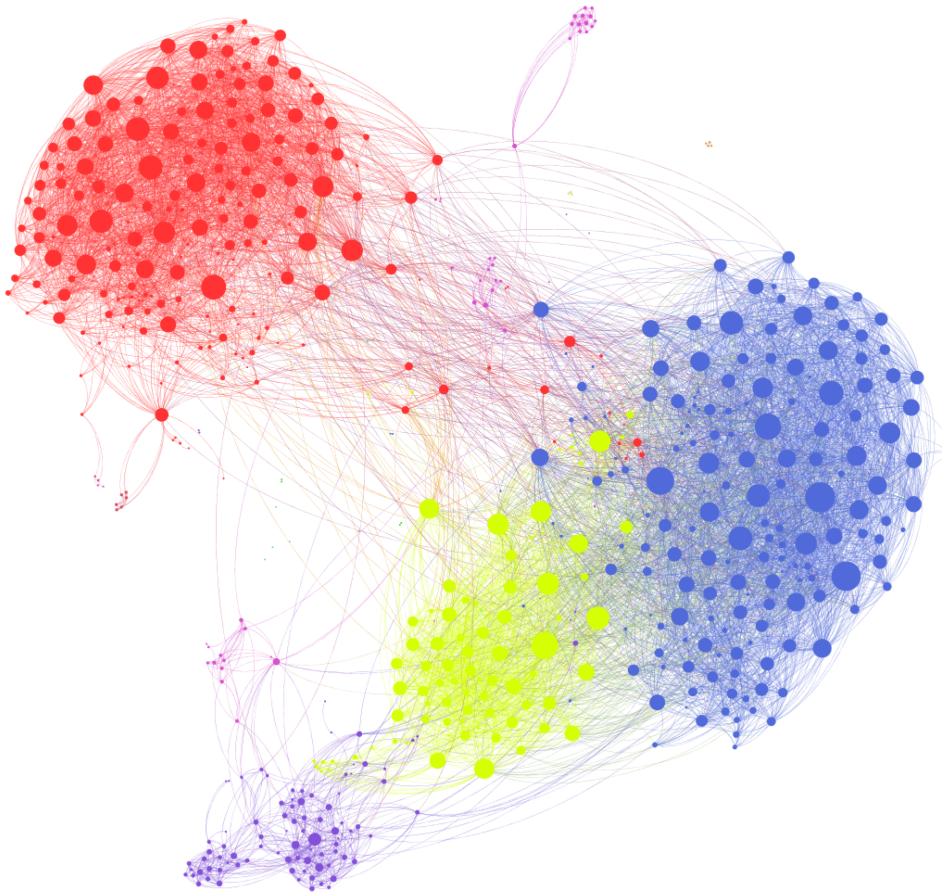
Michael B. Eisen,* Paul T. Spellman,* Patrick O. Brown,† and David Botstein*‡

Applications of clustering in ML

- Image segmentation: cluster similar regions of an image

Image: Jessica Wu

Applications of clustering in ML



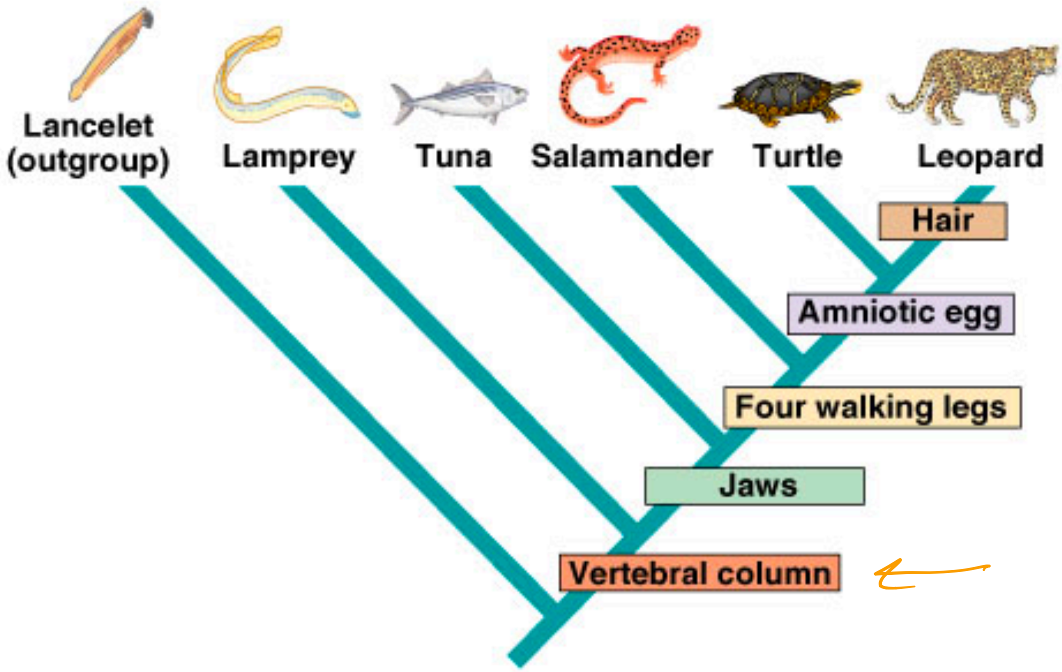
- Clustering in social graphs

Image: https://griffsgraphs.wordpress.com/2012/07/02/a-facebook-network/

Two main types of clustering

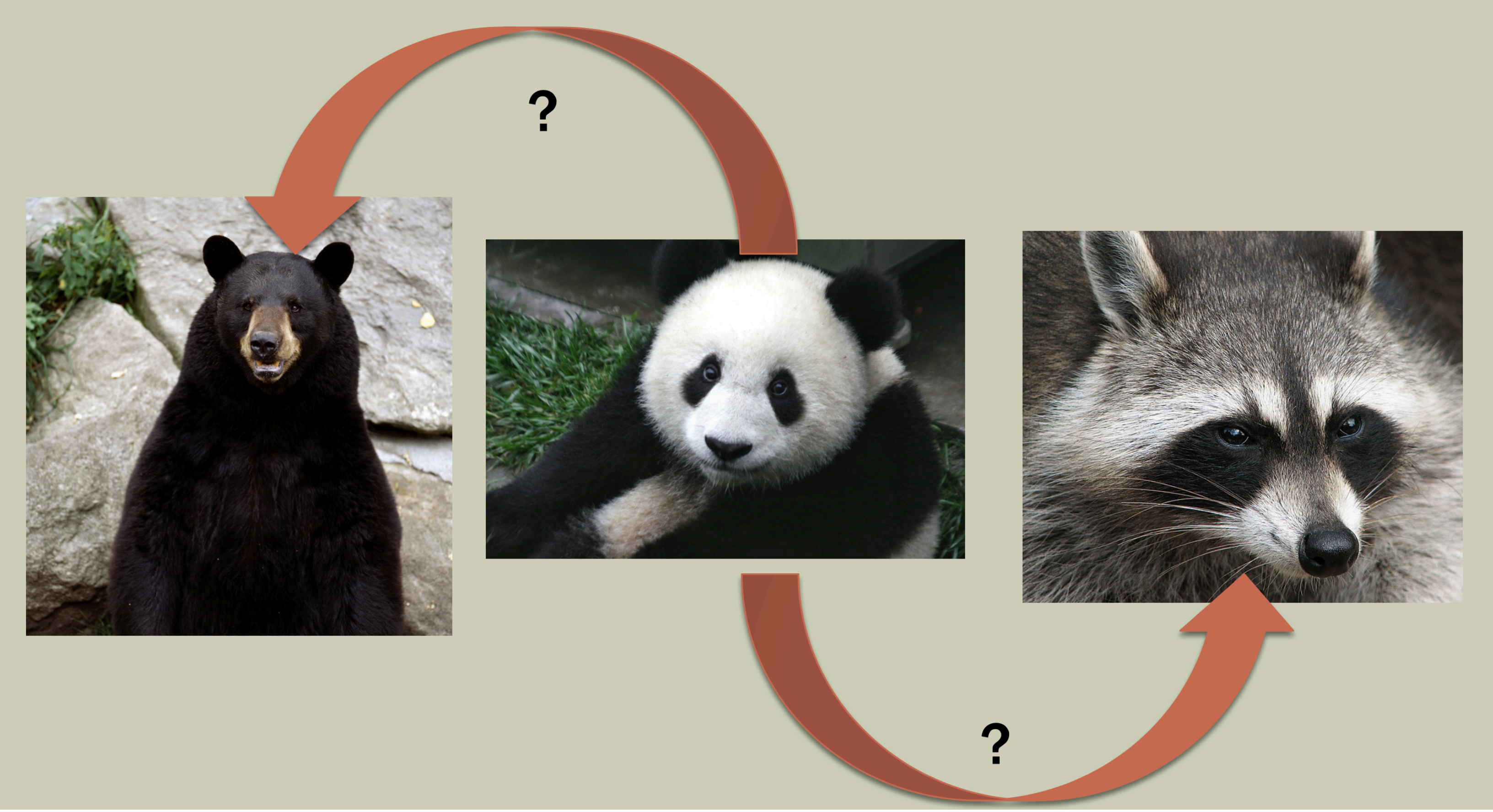
- Flat/Partitional:
  - K-means
  - Gaussian mixture models
- Hierarchical:
  - Agglomerative: bottom-up
  - Divisive: top-down
  - Examples: UPGMA and Neighbor Joining

Hierarchical clustering example: trees



Credit: Pearson Education, Benjamin Cummings

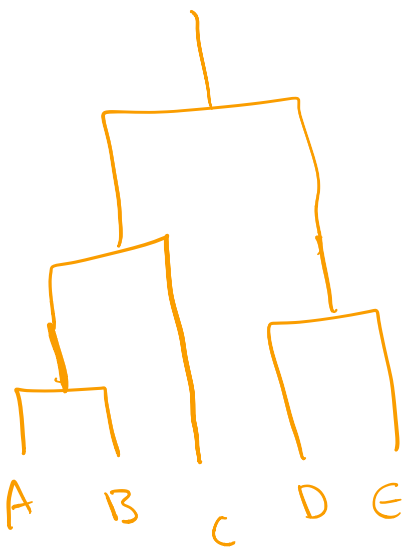
Are pandas more closely related to bears or raccoons?



Credit: Ameet Soni

UPGMA and Neighbor Joining

- Start with a dissimilarity map between examples (symmetric matrix)
- Say our examples are: A,B,C,D,E



δ	A	B	C	D	E
A	0	1	3	6	6
B		0	2	5	5
C			0	5	5
D				0	2
E					0

Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

University of Southampton
School of Biological Sciences

A D B F G C E

Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

UNIVERSITY OF Southampton
School of Biological Sciences

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

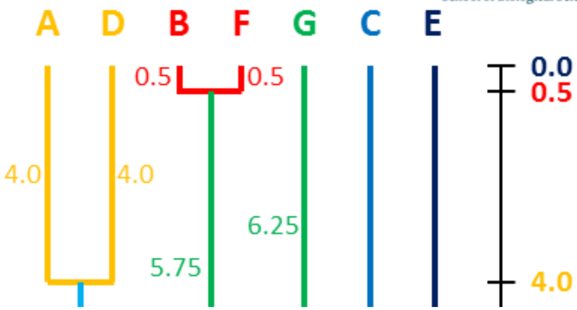


Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

UNIVERSITY OF Southampton
School of Biological Sciences

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

	AD	BFG	C
BFG	**16.50**		
C	26.50	30.67	
E	32.00	33.00	41.00

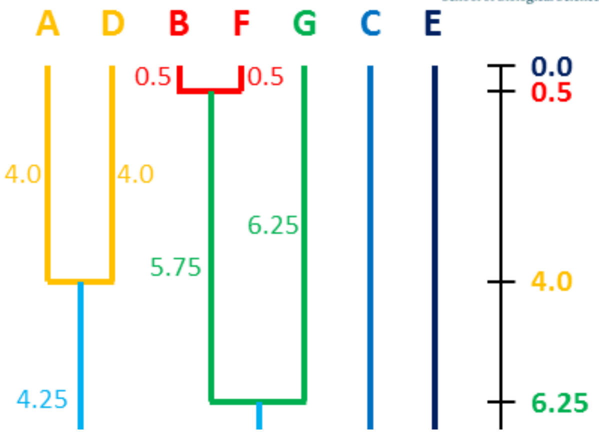


Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

University of Southampton
School of Biological Sciences

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

	AD	BFG	C
BFG	**16.50**		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	**29.00**	
E	32.60	41.00

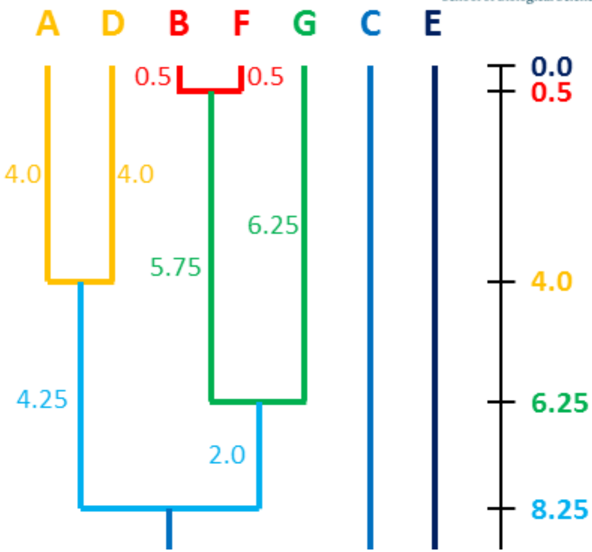


Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

UNIVERSITY OF Southampton
School of Biological Sciences

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

	AD	BFG	C
BFG	**16.50**		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	**29.00**	
E	32.60	41.00

	ADBFGC
E	**34.00**

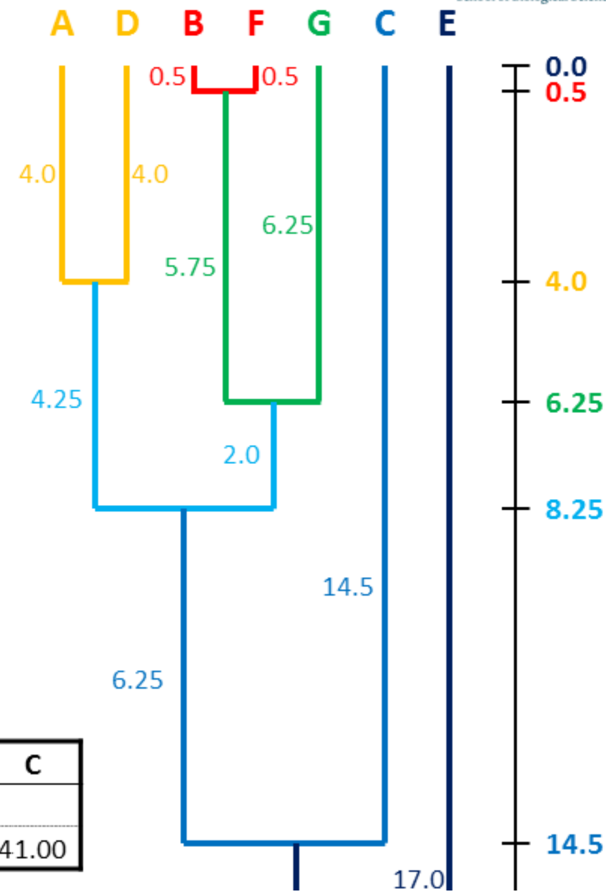


Figure: Dr. Richard Edwards

Hierarchical clustering example (UPGMA)

UNIVERSITY OF Southampton
School of Biological Sciences

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	**1.00**	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	**8.00**	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	**12.50**	29.00	28.00

	AD	BFG	C
BFG	**16.50**		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	**29.00**	
E	32.60	41.00

	ADBFGC
E	**34.00**

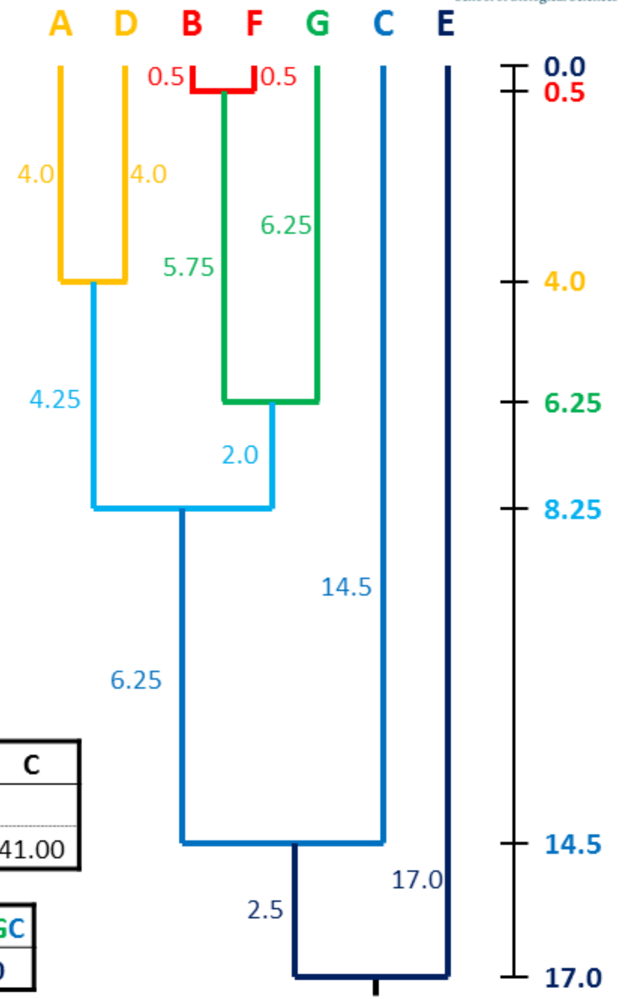
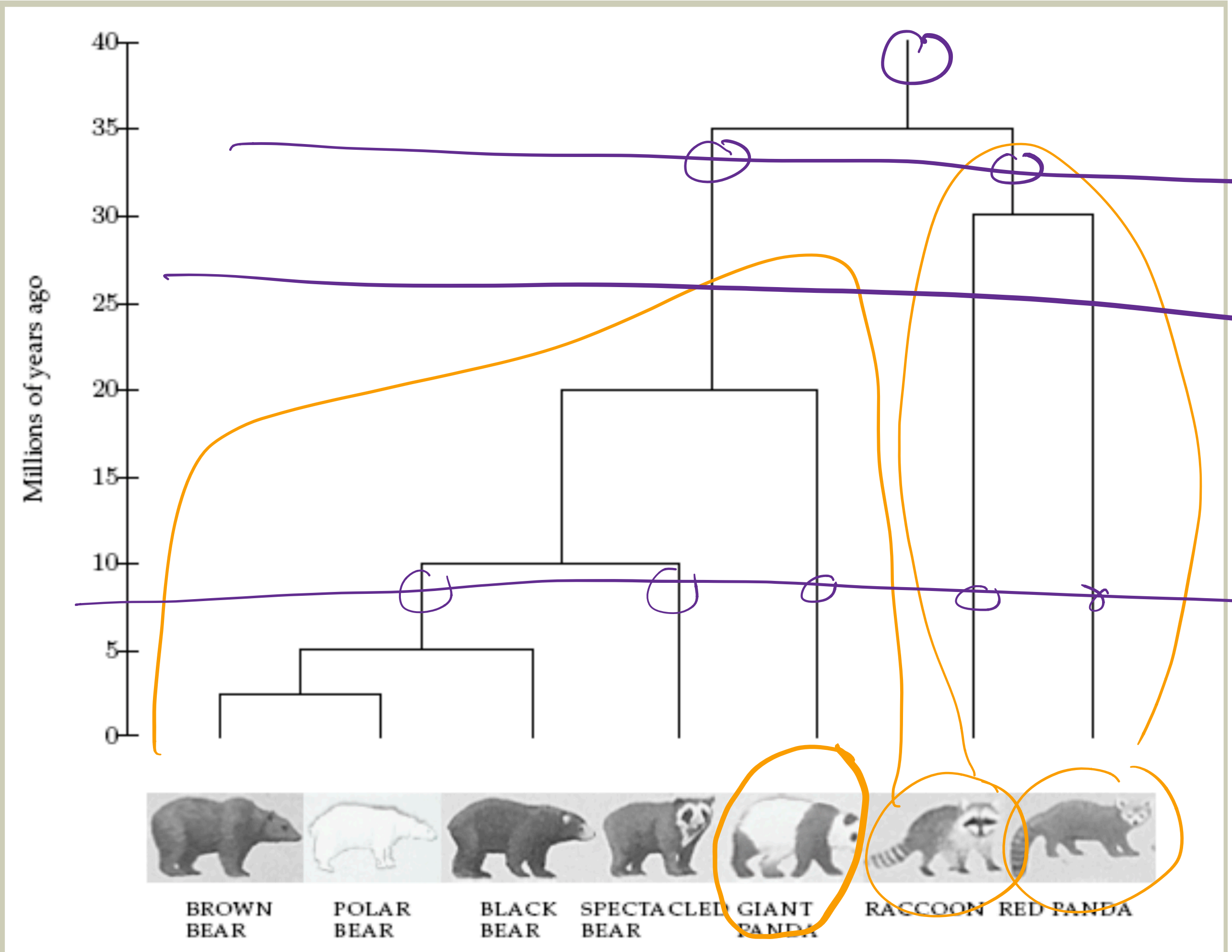


Figure: Dr. Richard Edwards

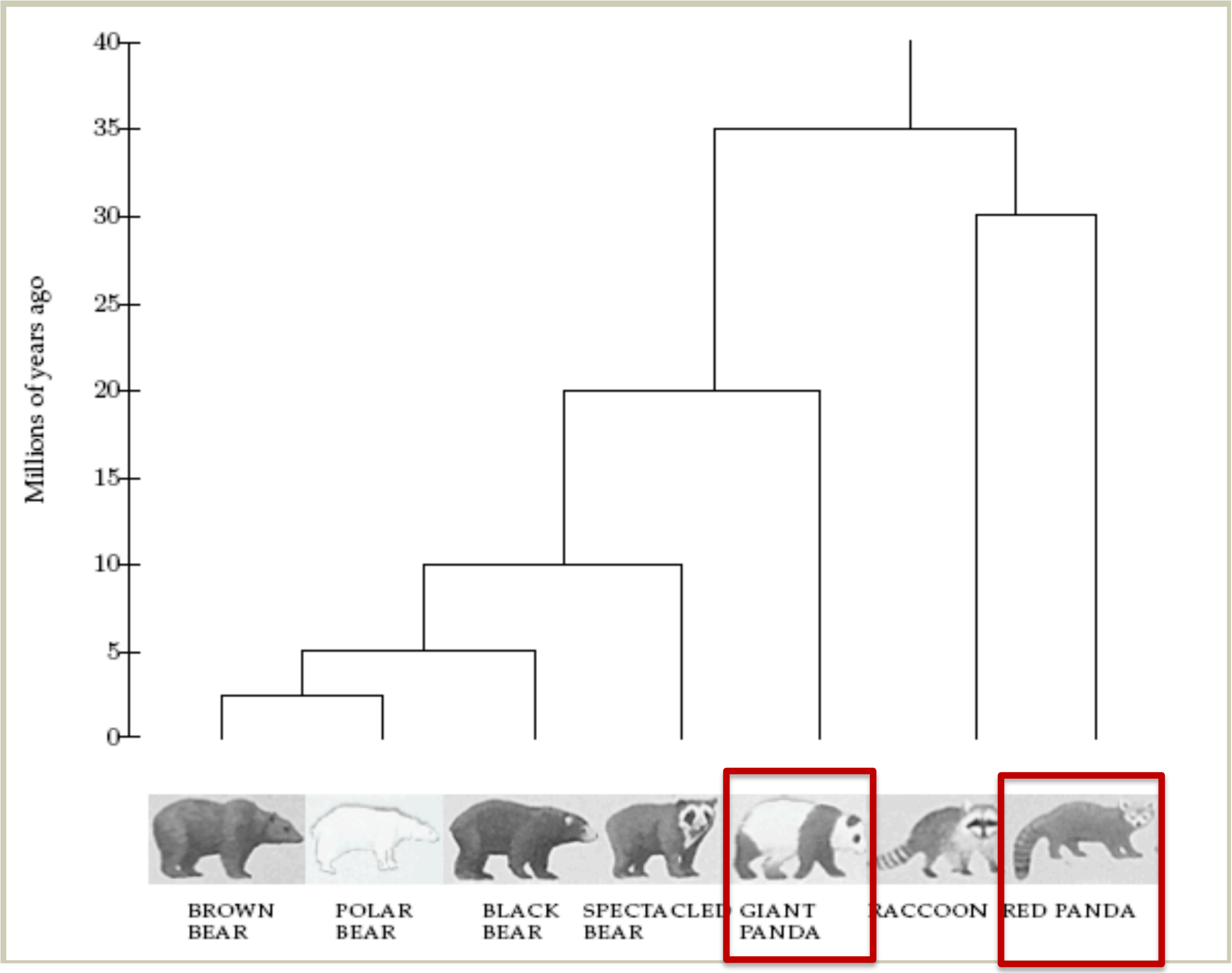
Back to the pandas….

Back to the pandas....



Credit: Ameet Soni

Back to the pandas….



Credit: Ameet Soni