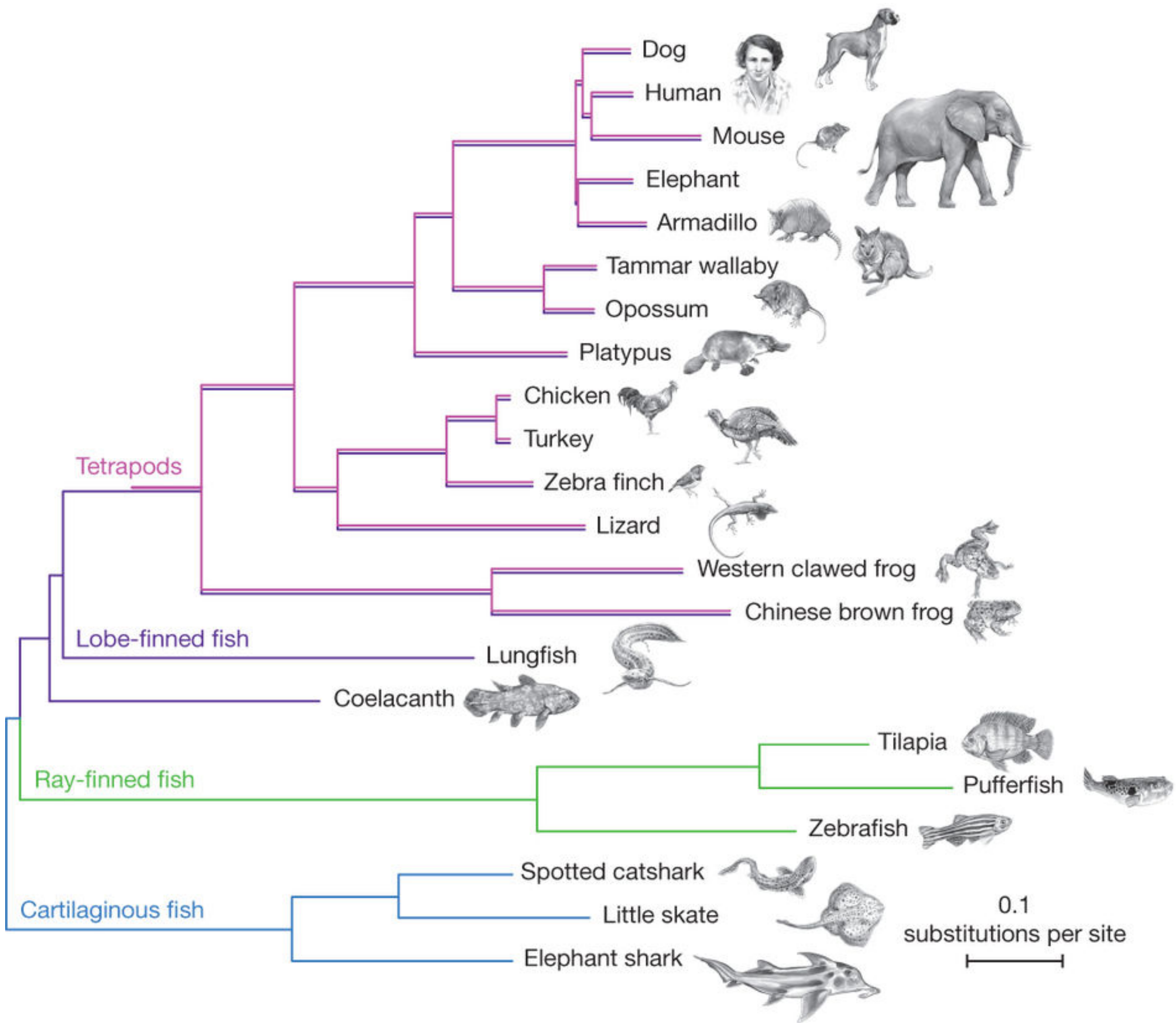
CS 189 Spring 2021 — Introduction to Machine Learning — Jonathan Shewchuk — DIS13

1 Hierarchical Clustering for Phylogenetic Trees

A phylogenetic tree (or "evolutionary tree") is way of representing the branching nature of evolution. Early branches represent major divergences in evolution (for example, modern vertebrae diverging from modern invertebrate), while later branches represent smaller branches in evolution (for example, modern humans diverging from modern monkeys). An example is shown below.



Creating phylogenetic trees is a popular problem in computational biology. We are going to combine what we know about clustering, decision trees, and unsupervised learning.

We start with all the samples (in this case, animals) in a single cluster and gradually divide this up, until each animal is in its own cluster. This should remind you of decision trees! After k steps, we have at most 2^k clusters. Since we do not have labels, we need to find some way of deciding how to split the samples (other than using entropy).

We will use the same objective as in k-means clustering to determine how good our proposed clustering is (here, $|x|$ denotes length and $||x||$ denotes norm):

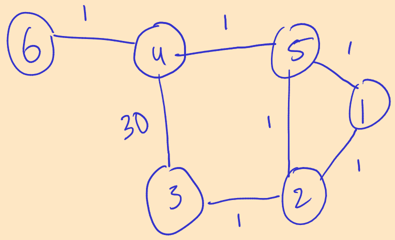
$$\min L = \sum_{i=1}^{k} \sum_{X_j \in S_i} ||X_j - \mu_i||_2^2, \text{ where}$$
$$\mu_i = \frac{1}{|S_i|} \sum_{X_j \in S_i} X_j, \quad i = 1, 2, \ldots, k.$$

At each iteration, we will split each cluster with more than one element into two clusters, choosing the split that achieves the minimum resulting objective value L. The algorithm terminates when every sample point is in its own cluster, yielding an objective value $L = 0$.

(a) Consider the following six animals and their two features. Create the resulting decision tree.

Animal	Lifespan	Wings
Dog	12	0
Human	80	0
Mouse	2	0
Elephant	60	0
Chicken	8	2
Turkey	10	2

(b) Prove that an optimal clustering on $k + 1 < n$ clusters has an objective value that is at least as small as that of the optimal clustering on k clusters.



Adjacency Matrix

A

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \hline \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 30 & 0 & 0 \\ 0 & 0 & 30 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

Degree Mat

D

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \hline \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 2 & & & & \text{O} & \\ & 3 & & & & \\ & & 2 & & & \\ & \text{O} & & 3 & & \\ & & & & 3 & \\ & & & & & 1 \end{bmatrix} \end{array}$$

Laplacian Matrix

$$L = D - A$$

$$\begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \hline \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

2 Spectral Clustering

In this question we will provide some intuition on spectral clustering in the context of undirected graphs. Consider a graph $G = (V, E)$ of n vertices and m edges. The adjacency matrix of a graph is $A \in \mathbb{R}^{n,n}$ matrix such that:

$$A_{ij} = \begin{cases} w_{ij} & (i,j) \in E \\ 0 & o.w. \end{cases}$$

where w_{ij} is the weight of the edge from i to j. Assume there are not self-edges, i.e. $(i,i) \notin E$, and that $\forall (i,j) \in E$, $w_{ij} \geq 0$

The Laplacian matrix of the graph G is defined as $L = D - A$, where D is a diagonal matrix whose entry D_{ii} is the sum of weights for all edges incident to vertex i, i.e. $D_{ii} = \sum_{j=1}^{n} w_{ij} \mathbb{1}[(i,j) \in E]$.

(a) Assume $w_{ij} = w_{ji}$ for all $(i,j) \in E$. Note that the following is an alternative way of writing the Laplacian L:

$$L_{ij} = \begin{cases} -w_{i,j} & (i,j) \in E \\ \sum_{j:(i,j)\in E} w_{i,j} & i = j \\ 0 & o.w. \end{cases}$$

Suppose that G is connected. Show that the all ones vector $\mathbf{1} \in \mathbb{R}^n$ is an eigenvector of L with eigenvalue 0.

$$(L\mathbb{1})_i = L_{i*}\mathbb{1} = \sum_{j=1}^{n} L_{ij} = \underbrace{\sum_{j:(i,j)\in E} w_{ij}}_{\text{diag}} - \underbrace{\sum_{j:(i,j)\in E} w_{ij}}_{\text{off-diag}} = 0$$

(b) Show that for any vector $x \in \mathbb{R}^n$, $x^\top L x = \sum_{(i,j)\in E} w_{ij}(x_i - x_j)^2$

$$x^\top L x = \sum_{i,j} L_{ij} x_i x_j = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij}\mathbb{1}\{(i,j)\in E\}\right)x_i^2 - \sum_{(i,j)\in E} w_{ij} x_i x_j$$

$$= \sum_{(i,j)\in E}\left[w_{ij}x_i^2 + w_{ij}x_j^2 - 2w_{ij}x_i x_j\right]$$

$$= \sum_{(i,j)\in E} \underbrace{w_{ij}}_{\geq 0}\underbrace{(x_i - x_j)^2}_{\geq 0}$$

$(i,j)\in E \Updownarrow (j,i)\in E$

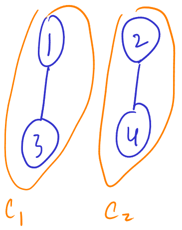
$w_{12}x_1x_2$
$w_{21}x_2x_1$

(c) Show that L is positive semi-definite.

$$x^\top L x \geq 0$$

(d) Assume G has k connected components. Construct k linearly independent eigenvectors in the nullspace of L.

$C_1, C_2, \ldots, C_k$. $\mathbb{1}_\ell[i] = \begin{cases} 1 & \text{if } i \in C_\ell \\ 0 & \text{o/w} \end{cases}$



$$\begin{array}{c|c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \end{array} \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}}_{\mathbb{1}_2}$$

$$\mathbb{1}_\ell^T L \mathbb{1}_\ell = \sum_{(i,j) \in E} w_{ij} \left(\mathbb{1}_\ell[i] - \mathbb{1}_\ell[j] \right)^2 = 0$$

But $L \succcurlyeq 0$.

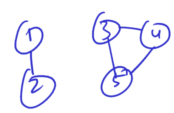
$\therefore \mathbb{1}_\ell \in \text{Null } L$.

(e) Show that the dimension of the nullspace(L), which is equal to the number of zero eigenvalues of L, equals the number of connected components of G. Let G have k CC's.

$\dim(\text{Null } L) \geq k$ $\quad \dim(\text{Null } L) \leq k$ $\quad \dim(\text{Null } L) = k$ $\blacksquare$

$$v \in \text{Null } L \Rightarrow Lv = 0 \Rightarrow v^T L v = 0 \Rightarrow \sum_{(i,j) \in E} w_{ij} (v_i - v_j)^2 = 0 \Rightarrow \left(w_{ij} \neq 0 \Rightarrow v_i = v_j \right)$$

$\Updownarrow$



$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

For every conn comp, v has a const val. $\mathbb{1}_\ell$

(f) Compose b linearly independent vectors $\{v_1, \ldots, v_b\}$ in the nullspace of L into a matrix $V =$ $\mathbb{1}_1 \cdots \mathbb{1}_b$ $[v_1, \ldots, v_b]$ (i.e. the vectors are the columns of V). Prove that if two rows of V, V_i and V_j, are equal, nodes i and j belong to the same connected component.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$\mathbb{1}_1 \quad \mathbb{1}_2$

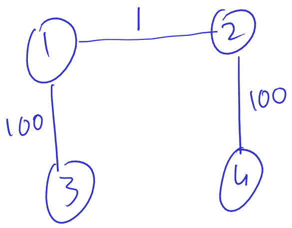
$$T \begin{cases} \mathcal{V} = \{v_1, \ldots, v_b\} \\ \mathcal{J} = \{\mathbb{1}_1, \ldots, \mathbb{1}_b\} \end{cases}$$

$V_i = V_j$
$\downarrow T^T$

(g) Explain how we could cluster the vertices of G into b clusters using V.

$V_i = V_j \Rightarrow i$ & j in the same cluster

(h) Suppose now we were to add edges with very small weights between the connected components of G, to make a fully connected graph G'. G' has only one eigenvector with eigenvalue zero (i.e. $\mathbf{1}$) so we choose the b eigenvectors of L with the smallest eigenvalues. Because the added weights were small, these will be relatively close to vectors in the null-space of L. Explain how we could construct b clusters now using the k-means algorithm.



$V_i' \approx V_j'$ if i & j are in CC.

k clusters.