

02-725 HW 1

Anushka Sinha

1.1

1. Starting from left to right, the number of hydrogens is 2, 1, 0, 3
2. Following the C,O,N,H,S order these are the explicit hydrogens: 23, 1, 0, 1, 0. The last hydrogen is attached the OH group which will give us 26 hydrogens in total.

1.2

1. Implicit adjacency matrix for Caffeine:

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

2. Explicit adjacency matrix for Caffeine:

[illegible]

3. Molecular formula (consider atoms C, O, N, H and S):

Penicillin A
 $\begin{bmatrix} 16 & 4 & 2 & 26 & 1 \end{bmatrix}$

Penicillin G
 $\begin{bmatrix} 16 & 4 & 2 & 17 & 1 \end{bmatrix}$

Caffeine
 $\begin{bmatrix} 8 & 2 & 4 & 10 & 0 \end{bmatrix}$

Label paired with length = 3

Penicillin A
 $\begin{bmatrix} 220 & 20 & 34 & 144 & 14 \\ 20 & 0 & 3 & 9 & 0 \\ 34 & 3 & 4 & 8 & 2 \\ 144 & 9 & 8 & 56 & 8 \\ 14 & 0 & 2 & 8 & 0 \end{bmatrix}$

Penicillin G
 $\begin{bmatrix} 218 & 19 & 34 & 99 & 14 \\ 19 & 0 & 3 & 6 & 0 \\ 34 & 3 & 4 & 8 & 2 \\ 99 & 6 & 8 & 12 & 8 \\ 14 & 0 & 2 & 8 & 0 \end{bmatrix}$

Caffeine
 $\begin{bmatrix} 38 & 13 & 69 & 60 & 0 \\ 13 & 0 & 1 & 0 & 0 \\ 69 & 1 & 6 & 0 & 0 \\ 60 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Depth first search (all cycles, double traverse, no compression, depth = 2). This should be a binary (0/1)

map

Penicillin A

1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0,
 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0

Penicillin G

1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0,
 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0

Caffeine

1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0,
 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0

4. Using your mappings compute the following kernels $k(G1, G2)$ for all $\binom{3}{2}$ pairs of Penicillin A, G, and Caffeine:

Molecular formula

$k(G1, G2)$ using the molecular formula mapping for Penicillin A and Caffeine is 404

$k(G1, G2)$ using the molecular formula mapping for Penicillin G and Caffeine is 314

$k(G1, G2)$ using the molecular formula mapping for Penicillin A and Penicillin G is 719

Molecular formula + MinMax

Molecular Formula + Min Max $k(G1, G2)$ for Penicillin A and Caffeine is 0.43

Molecular Formula + Min Max $k(G1, G2)$ for Penicillin G and Caffeine is 0.52

Molecular Formula + Min Max $k(G1,G2)$ for Pennicilin A and Pennicilin G is 0.82

Label paired

The Label paired $k(G1,G2)$ for Penicillin A and Caffeine is 30882

The Label paired $k(G1,G2)$ for Penicillin G and Caffeine is 25380

The Label paired $k(G1,G2)$ for Penicillin A and Penicillin G is 81014

Label paired + MinMax

Label Pair + Min Max $k(G1,G2)$ for Pennicilin A and Caffeine is 0.309

Label Pair + Min Max $k(G1,G2)$ for Pennicilin G and Caffeine is 0.373

Label Pair + Min Max $k(G1,G2)$ for Pennicilin A and Pennicilin G is 0.812

Depth first search

Depth First Search $k(G1,G2)$ for Pennicilin A and Caffeine is 22

Depth First Search $k(G1,G2)$ for Pennicilin G and Caffeine is 22

Depth First Search $k(G1,G2)$ for Pennicilin A and Pennicilin G is 40

Depth first search + Tanimoto

DFS + Tanimoto Similarity $k(G1,G2)$ for Pennicilin A and Caffeine is 0.4783

DFS + Tanimoto Similarity $k(G1,G2)$ for Pennicilin G and Caffeine is 0.55

DFS + Tanimoto Similarity $k(G1,G2)$ for Pennicilin A and Pennicilin G is 0.869

2

Let feature map for k_1 be given as $\phi_1 : X \rightarrow \mathbb{R}^n$ and feature map for k_2 be given as $\phi_2 : X \rightarrow \mathbb{R}^m$.

Let $\phi : X \rightarrow \mathbb{R}^{n+m}$ be the feature map given by the formula $\phi(x) = (\phi_1(x), \phi_2(x))$. The kernel(k) for this feature map is thus given by $k(x, y) = \phi(x)^T \phi(y) = \phi_1(x)^T \phi_1(y) + \phi_2(x)^T \phi_2(y) = k_1(x, y) + k_2(x, y)$ as asked.