School of Computing and Information Systems
The University of Melbourne
COMP90049 Knowledge Technologies (Semester 2, 2019)
Workshop exercises: Week 3

1. Finish any remaining questions from last week, if necessary.

---

2. Consider the following collection of "documents" C:

   (i) *It is what it is.*

   (ii) *Jean's hat is finer than Karl's hat.*

   (iii) *We are obsessing about gene issues.*

   Build a feature vector for all the documents in this collection.

3. Based on the following metrics decide which of the "documents" in C is most similar to (*iv*)
   (*iv*) *Karl is obsessed with genes.*

   (a) Euclidean distance

   (b) Cosine similarity

   (c) Jaccard similarity

4. Please explain the difference between *Distance* and *Similarity* calculations in question 3.

---

5. Consider we have two coins one *fair* coin and another one with two *heads*. Both coins are in a bag. In our test trial, we randomly select a coin from the bag, toss it and check the results.

   (a) Using the Bayes Rules, calculate the prior and posterior probability of choosing the fair coin in this experiment. (In your calculation, consider both possibilities of observing a *head* or a *tail* in our test trial).

   (b) We repeat the experience again. Calculate the probability that we have chosen the fair coin (in both experiments) if we observe two *Heads* in a row.

6. Calculate the **entropy** for our first trial in question 5.