HDDA Tutorial: Distance : Solutions

Department of Econometrics and Business Statistics, Monash University

Tutorial 3

Work in groups of two people:

1. Consider the *age* (in years) and *height* (in cm) of both you and the other person (you are allowed to lie about these). Compute the Euclidean distance between you and the other person for these two variables.

Euclidean distance between National Basketball Association (NBA) most valuable player Giannis Antentok-ounmpo (height 211cm, age 24 years) and footballer Leo Messi (height 170 cm, age 32) is 41.77.

2. Repeat question 1 but use the Manhattan distance.

Manhattan distance between National Basketball Association (NBA) most valuable player Giannis Antentok-ounmpo and footballer Leo Messi is 41+8=49.

3. Repeat question 2 but measure height in metres.

The Manhattan distance is now 0.41+8=8.41. The units of measurement influence distance. This is why data in different units are normally standardised before computing distances. If both variables are measured in the same units (e.g. height and length of hands both measured in cm) then this standardisation is not necessary.

Select from the following list the types of cuisines that you enjoy:

- Chinese food
- Indian food
- Italian food
- Japanese food
- Lebanese food
- Mexican food
- Thai food
- British food

4. Compute a Jaccard similarity between you and the other person with regards to your taste in food.

Consider someone who only likes Chinese food and someone who likes Chinese Thai, Italian and Japanese. In common they enjoy one cuisine (Chinese). Between them they enjoy four cuisines (Chinese, Thai, Italian and Japanese). The Jaccard similarity is therefore 1/4 or 0.25.

5. Compute a Jaccard distance between you and the other person with regards to your taste in food.

Jaccard distance is 1 minus Jaccard similarity. For the previous example Jaccard distance is 0.75.

6. How would you define a distance between you and the other person that takes into account height, age and food preference.

One idea may be to add the Manhattan (or Euclidean) Distance and Jaccard similarity together. This will also satisfy the axioms of a distance (non-negativity, identity of indiscernables, symmetry, triangle inequality). As an advanced exercise keen student can try to prove that if two distances respects the four axioms their sum will also respect the four axioms.