HDDA Tutorial: Distance : Solutions

Department of Econometrics and Business Statistics, Monash University

Tutorial 3

Work in groups of two people:

1. Consider the *age* and *height* 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 Antentokounmpo (height XXcm, age XX) and European Footballer of the year Luka Modric (height XX cm, age XX) is XX. Notice that units of measurement affect calculation.

2. Repeat question 1 but use the Manhattan distance.

Manhattan distance between National Basketball Association (NBA) most valuable player Giannis Antentokounmpo (height XXcm, age XX) and European Footballer of the year Luka Modric (height XX cm, age XX) is XX. Notice that units of measurement affect calculation.

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
```

3. 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.

4. 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.

5. 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.