K-neareast Neighbors

Instruction: Following the sample problems and solution located in the end of this AYU to do this AYU.

### Problem 1.

Given the data.

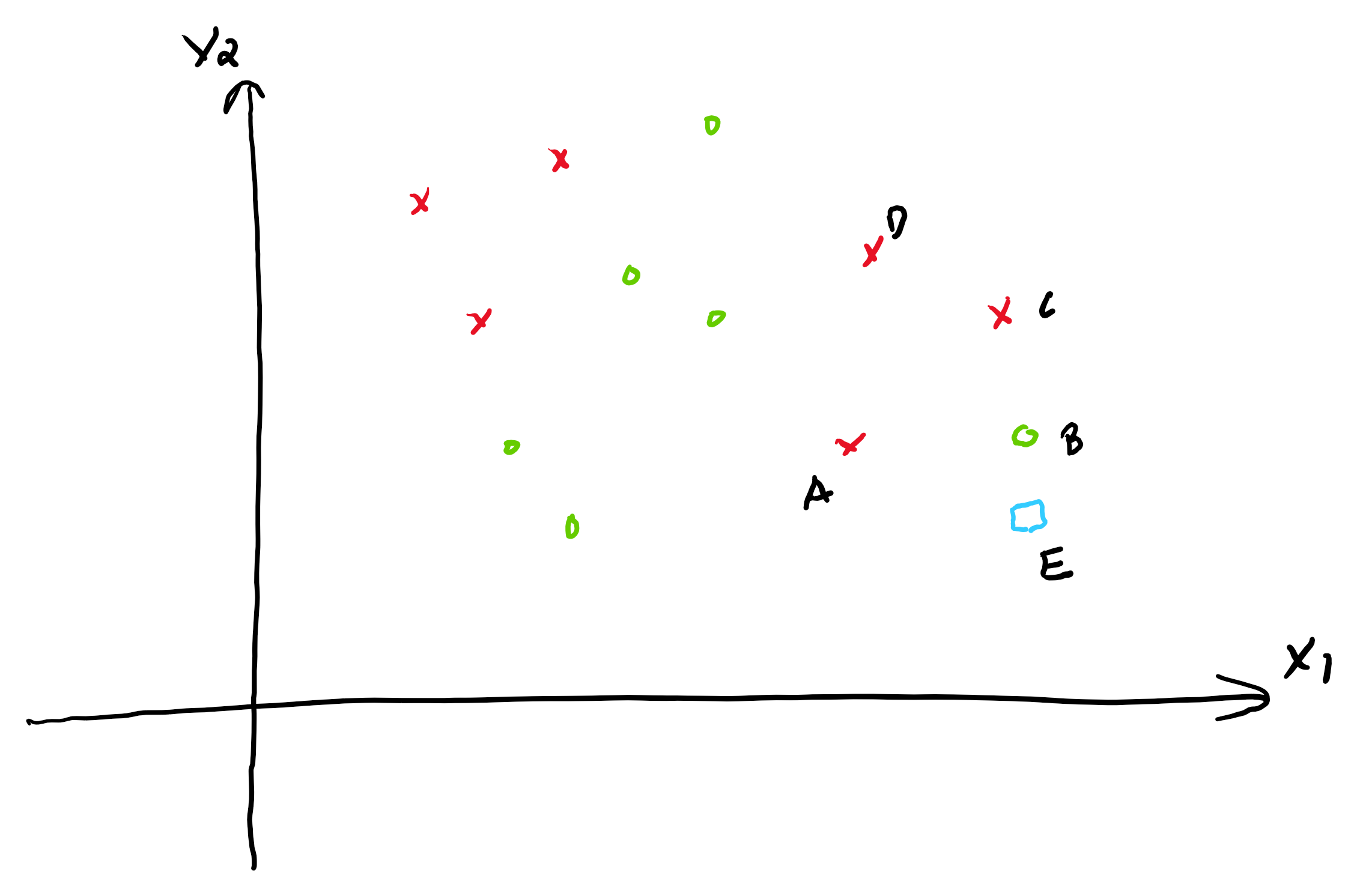
|  | Age | Sex | Survived |
| --- | --- | --- | --- |
| A | 27 | M | 0 |
| B | 30 | F | 1 |
| C | 80 | F | 1 |
| D | 50 | M | 0 |
| E | 60 | F | 0 |
| F | 70 | F | 1 |

1. Let G be a female of 55 years old. Use 1NN to predict whether G is survived (Survived =1) or not (Survived = 0). Does the prediction change if used 3NN?
2. Given the following data, use 1NN and 3NN to predict the salary for G (a female of 55 years old).

|  | Age | Sex | Salary (k) |
| --- | --- | --- | --- |
| A | 27 | M | 80 |
| B | 30 | F | 70 |
| C | 80 | F | 90 |
| D | 50 | M | 60 |
| E | 60 | F | 10 |
| F | 70 | F | 100 |

### Problem 2.

Given the data. Consider as 1 and as 0.



With EB = 1.4, EA = 3, EC = 3, ED = 4,

1. Use the **uniform weights** to calculate the predicted probability and the prediction of **3NN** for E.
2. Use the **distance weights** to calculate the predicted probability and the prediction of **3NN** for E.
3. Use the **distance weights** to calculate the predicted probability and the prediction of **4NN** for E.

### Problem 3 (Optional)

We can use KNN to build a recommendation system to recommend items to users.

[KNN for Recommendation System](ayu_7_rs.html)

Given the utility matrix

|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| --- | --- | --- | --- | --- | --- |
| Alice | 5 | 3 | 3 | 4 |  |
| User 1 | 3 | 1 | 2 | 3 | 3 |
| User 2 | 2 | 3 | 4 | 3 | 5 |
| User 3 | 3 | 3 | 1 | 4 | 4 |
| User 4 | 1 | 5 | 5 | 4 | 2 |

Should we recommend Item 5 to Alice? Calculate her estimated rating on Item 5 to answer the question. Recommend the item if Alice’s rating is 4 or above.

1. Use user-based KNN, with and Manhantan distance.
2. Use item-based KNN, with and cosine similarity.

–

## Sample Problems and Solutions

### Problem 1

Given the data.

|  | Age | Sex | Survived |
| --- | --- | --- | --- |
| A | 27 | M | 0 |
| B | 30 | F | 1 |
| C | 80 | F | 1 |
| D | 50 | M | 0 |
| E | 60 | F | 0 |

1. Let F be a female of 40 years old. Use 1NN to predict whether F is survived (Survived =1) or not (Survived = 0). Does the prediction change if used 3NN
2. Given the following data, use 1NN and 3NN to predict the salary for F.

|  | Age | Sex | Salary ($1000) |
| --- | --- | --- | --- |
| A | 27 | M | 80 |
| B | 30 | F | 70 |
| C | 80 | F | 90 |
| D | 50 | M | 60 |
| E | 60 | F | 10 |

### Solution

*Step 1:* Standardize the Age column.

We have the follows.

|  | Age | Sex | Survived |
| --- | --- | --- | --- |
| A | 0 | M | 0 |
| B | 0.057 | F | 1 |
| C | 1 | F | 1 |
| D | 0.434 | M | 0 |
| E | 0.623 | F | 0 |
| F | 0.245 | F |  |

*Step 2:* Calculate the distances from F to all the samples.

We have

We have BF < EF < CF < DF < AF. B is the nearest neighbor and B, E, C are the three nearest neighbor.

Since the nearest neighbor of F is B (BF is the smallest between AF, BF, CF, DF, and EF), thus we predicts F has the same outcome as B, which is Survived.

If we use 3NN, the prediction of F would be the majority outcome of the 3 nearest neighbors. The three nearest neighbor of F are B, E and C. The majority outcomes of B, E and C is still Survived. Thus, 3NN also predicts F Survived. The prediction does not change from 1NN to 3NN in this case.

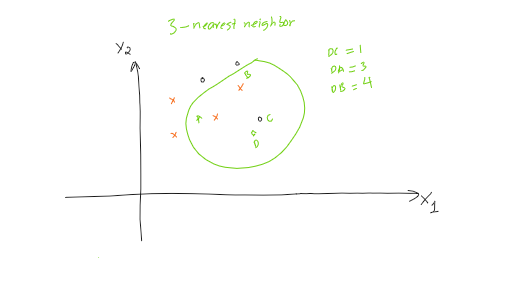
1. Since the distances does not change, the neareast neighbors are still the same as part a.

For 1NN, the predicted salary of F is the same as B, 70k.

For 3NN, the predicted salary of F is the average salaries of B, E and C, which is:

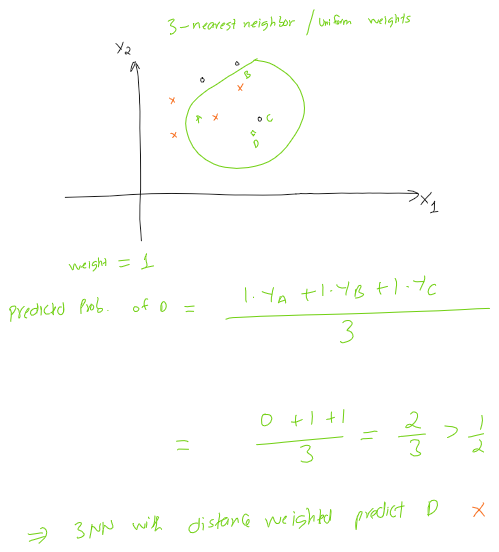
### Problem 2

Given the data. Consider as 1 and as 0.

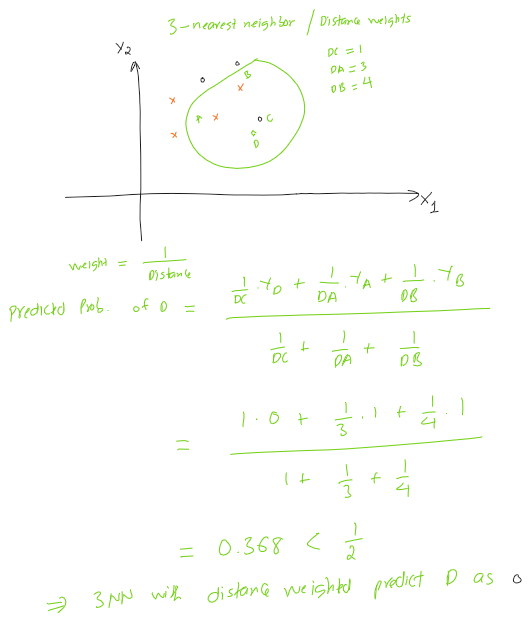


1. Use the **uniform weights** to calculate the predicted probability and the prediction of **3NN** for D.
2. Use the **distance weights** to calculate the predicted probability and the prediction of **3NN** for D.

### Solution

1. 

Notice that if the probability is greater or equal to 1/2, we predict x.

1. 

### Problem 3

Given the utility matrix

|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
| --- | --- | --- | --- | --- | --- |
| Alice | 5 | 3 | 4 | 4 |  |
| User 1 | 3 | 1 | 2 | 3 | 3 |
| User 2 | 4 | 3 | 4 | 3 | 5 |
| User 3 | 3 | 3 | 1 | 4 | 4 |
| User 4 | 1 | 5 | 5 | 2 | 1 |

Should we recommend Item 5 to Alice? Calculate her estimated rating on Item 5 to answer the question. Recommend the item if Alice’s rating is 4 or above.

1. Use user-based KNN, with and Manhantan distance.
2. Use item-based KNN, with and cosine similarity.

### Solution

1. We calculate the Manhantance distance from Alice to other users. Let be the Manhantan distance from Alice to User 1 (exclude Item 5). We have.

Similarly,

$$ d\_{A4} = |5-1|+|3-5|+|4-5|+|4-2| = 9. \\
$$

Thus, the two nearest neighbors are User 2 and User 3. Alice’s rating on Item 5 is predicted as the average ratings of these two Users on Item 5, which is (5+4)/2= 4.5.

We recommend this Item to Alice.

1. We calculate the cosine similarities of Item 5 to other items (excluding Alice). Let be the cosine similarity of Item 5 and Item 1. We have.

Similarly,

Since and are the largest, the two nearest neighbors (largest similarities) of Item 5 are Item 1 and Item 4. Alice’s rating on Item 5 is predicted as the average ratings of her ratings for Item 1 and Item 4, which is (5+4)/2= 4.5.

We also recommend this Item to Alice.