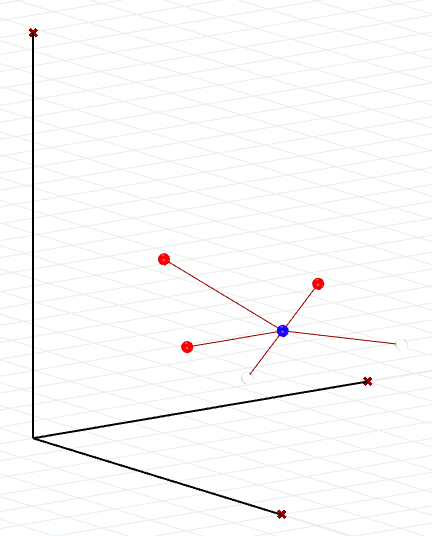
1. K-Nearest Neighbors

dataset = np.array([

[1, 2, 3, 0],

[2, 3, 1, 1],

[3, 1, 2, 0],

[4, 5, 1, 1],

[3, 3, 4, 0]

])

(a) **Distance Calculation**

The nearest 3 neighbors for K=3:

Neighbor Index: 1, Distance: 1.4142, Class Label: 1

Neighbor Index: 2, Distance: 2.0000, Class Label: 0

Neighbor Index: 4, Distance: 2.0000, Class Label: 0

Assigned class for K=3: 0

(b) **Impact of K**

Assigned class for K=1: 1

Assigned class for K=5: 0

Small K Value:

-> Benefits: (1) Very sensitive the local variation of the data set;

1. Can deal with the situation where there are many small group of data.

-> Drawbacks: (1) over-fitting, which means the model will be specific sensitive the training data and can not fit the new data very well;

1. Will be disturbed by abnormal data, if the new observation is surrounded by some wrong data, the model will make the wrong decision;
2. The model can not have a very good „understanding“ of the global data set.

Big K Value:

-> Benefits: (1) Will not be disturbed by wrong data, when there is a relative big data set, big K value will lead to a more average result;

(2) Have a better understanding of the whole data set.

-> Drawbacks: (1) lack of precision, which means some small group will be ignored;

1. Oversimplified.

(c) **Distance Weighting**

Some type of distance-weighted voting:

**Inverse Distance Weighting** - In this type of voting, the closer the data is, the more impossible for the new observation to be the same. It will r**everse the result**, which means when K=3, the new observation(X1=3, X2=3, X3=2) will be classified as **Class Label: 1**.

**Gaussian Weighting** - In this type, the weight will be calculated based on the distance using a Gaussian function like .So when K=3, the nearest three neighbors are at indices 1, 2, and 4 in the data set, their new distance will be:

Neighbor 2 (index 1): New\_distance = 1.4142 \* 0.1353 = 0.1913

Neighbor 3 (index 2): New\_distance = 2 \* 0.0183 = 0.1353

Neighbor 5 (index 4): New\_distance = 2 \* 0.0183 = 0.1353

New class label will 1.

**Code**

