# Machine Learning In Action — Chapter 2

## Example 2.1— Explaining kNN

In this example, the parameters are

$$k = 3, \qquad in\_X = \begin{pmatrix} 0.6 & 0.6 \end{pmatrix},$$

$$data\_set = \begin{pmatrix} 1.0 & 1.1 \\ 1.0 & 1.0 \\ 0.0 & 0.0 \\ 0.0 & 0.1 \end{pmatrix}, \qquad labels = \begin{pmatrix} A \\ A \\ B \\ B \end{pmatrix}$$

Using the above arguments,

$$\begin{aligned} diff\_matrix &= \textbf{tile}\left(in\_X, \mathbf{(4,\ 1)}\right) - data\_set \\ &= \begin{pmatrix} 0.6 & 0.6 \\ 0.6 & 0.6 \\ 0.6 & 0.6 \\ 0.6 & 0.6 \end{pmatrix} - \begin{pmatrix} 1.0 & 1.1 \\ 1.0 & 1.0 \\ 0.0 & 0.0 \\ 0.0 & 0.1 \end{pmatrix} \\ &= \begin{pmatrix} -0.4 & -0.5 \\ -0.4 & -0.4 \\ 0.6 & 0.6 \\ 0.6 & 0.5 \end{pmatrix} \end{aligned}$$

and $diff\_matrix$ represent the straight line distance the vector is from the other data points in the set. To calculate the Euclidean Distance, take the square of every element in $diff\_matrix$, sum them across the rows, then take the square root.

$$sq\_diff\_matrix = \begin{pmatrix} 0.16 & 0.25 \\ 0.16 & 0.16 \\ 0.36 & 0.36 \\ 0.36 & 0.25 \end{pmatrix}, sq\_distances = \begin{pmatrix} 0.41 \\ 0.32 \\ 0.72 \\ 0.61 \end{pmatrix}$$

$sq\_distances$ represents the sum of all the distances from the vector is from the data point. Finally, taking square root of the distances and then sorting them,

$$distances = \begin{pmatrix} 0.64031 \\ 0.56568 \\ 0.84852 \\ 0.78102 \end{pmatrix}, sorted\_dist\_indices = \begin{pmatrix} 1 \\ 0 \\ 3 \\ 2 \end{pmatrix}$$

Taking the first 3 labels, we can see that there are 2 data points labelled $A$ and 1 data point labelled $B$. Hence, by majority, the data point $in\_X$ is classified in group $A$.