**Dina Ladnyuk**

**(A)** Normalise the data using Z-score normalisation.

z\_i = (xi-x\_mean)/sigma

**sepal length** x\_mean =5.79

i = 1 x1 - x\_mean = 6.3 - 5.79 = 0.51 (x1 - x\_mean)^2 = 0.26

i = 2 x2 - x\_mean = 5.1 – 5.79 = - 0.69 (x2 - x\_mean)^2 = 0.48

i = 3 x3 - x\_mean = 5.7 - 5.79 = - 0.09 (x3 - x\_mean)^2 = 0.01

i = 4 x4 - x\_mean = 5.0 - 5.79 = - 0.79 (x4 - x\_mean)^2 = 0.62

i = 5 x5 – x\_mean = 4.8 – 5.79 = -0.99 (x5 – x\_mean)^2 = 0.98

i = 6 x6 – x\_mean = 6.6 – 5.79 = 0.81 (x6 – x\_mean)^2 = 0.66

i = 7 x7 – x\_mean = 6.3 – 5.79 = 0.51 (x7 – x\_mean)^2 = 0.26

i = 8 x8 - x\_mean = 6.3 – 5.79 = 0.51 (x8 - x\_mean)^2 = 0.26

i = 9 x9 – x\_mean = 6.0 – 5.79 = 0.21 (x9 – x\_mean)^2 = 0.04

i = 10 x10 – x\_mean = 4.7 – 5.79 = -1.09 (x10 – x\_mean)^2 = 1.19

i=11 x11 - x\_mean = 5.0- 5.79 = -0.79 (x11 - x\_mean)^2 = 0.62

i=12 x12 – x\_mean = 6.1 - 5.79 = 0.31 (x12 – x\_mean)^2 = 0.10

i=13 x13 – x\_mean = 6.8- 5.79 = 1.01 (x13 – x\_mean)^2 = 1.02

i=14 x14 – x\_mean = 4.5 – 5.79 = - 1.29 (x14 – x\_mean)^2 = 1.66

i = 15 x15 – x\_mean = 7.7 – 5.79 = 1.91 (x15 – x\_mean)^2 = 3.64

sum((xi – x\_mean)^2) = 11.7

sigma^2 = sum((xi – x\_mean)^2) /15 = 0.78

sigma = 0.88

**z\_i = new x\_i = (xi-x\_mean)/sigma**

**sepal length:**

new x\_1 = (x1-x\_mean)/sigma = 0.51/0.88 = 0.58

new x\_2 = (x2 - x\_mean)/sigma = - 0.69/0.88 = -0.78

new x\_3 = (x3 - x\_mean)/sigma = - 0.09/0.88 = - 0.10

new x\_4 = (x4 - x\_mean)/sigma = - 0.79/0.88 = - 0.9

new x\_5 = (x5 - x\_mean)/sigma = - 0.99/0.88 = - 1.12

new x\_6 = (x6 - x\_mean)/sigma = 0.81/0.88 = 0.92

new x\_7 = (x7 - x\_mean)/sigma = 0.51/0.88 = 0.58

new x\_8 = (x8 - x\_mean)/sigma = 0.51/0.88 = 0.58

new x\_9 = (x9 - x\_mean)/sigma = 0.21/0.88 = 0.24

new x\_10 = (x10 - x\_mean)/sigma = - 1.09/0.88 = -1.24

new x\_11 = (x11 - x\_mean)/sigma = - 0.79/0.88 = -0.9

new x\_12 = (x12 - x\_mean)/sigma = 0.31/0.88 = 0.35

new x\_13 = (x13 - x\_mean)/sigma = 1.01/0.88 = 1.14

new x\_14 = (x14 - x\_mean)/sigma = -1.29/0.88 = -1.47

new x\_15 = (x15 - x\_mean)/sigma = 1.91/0.88 = 2.17

**sepal width:** x\_mean = 3.11

i = 1 x1 - x\_mean = 2.9 - 3.11= -0.21 (x1 - x\_mean)^2 = 0.04

i = 2 x2 - x\_mean = 3.4 – 3.11= 0.29 (x2 - x\_mean)^2 = 0.08

i = 3 x3 - x\_mean = 2.5 - 3.11 = - 0.61 (x3 - x\_mean)^2 = 0.37

i = 4 x4 - x\_mean = 3.5 - 3.11 = 0.39 (x4 - x\_mean)^2 = 0.15

i = 5 x5 – x\_mean = 3.4 – 3.11= 0.29 (x5 – x\_mean)^2 = 0.08

i = 6 x6 – x\_mean = 2.9 – 3.11= -0.21 (x6 – x\_mean)^2 = 0.04

i = 7 x7 – x\_mean = 3.3 – 3.11= 0.19 (x7 – x\_mean)^2 = 0.04

i = 8 x8 - x\_mean = 3.4 – 3.11= 0.29 (x8 - x\_mean)^2 = 0.08

i = 9 x9 – x\_mean = 3.4 – 3.11= 0.29 (x9 – x\_mean)^2 = 0.08

i = 10 x10 – x\_mean = 3.2 – 3.11= 0.09 (x10 – x\_mean)^2 = 0.01

i=11 x11 - x\_mean = 3.3 - 3.11 = 0.19 (x11 - x\_mean)^2 = 0.04

i=12 x12 – x\_mean = 2.9 - 3.11 = -0.21 (x12 – x\_mean)^2 = 0.04

i=13 x13 – x\_mean = 3.2- 3.11 = 0.09 (x13 – x\_mean)^2 = 0.01

i=14 x14 – x\_mean = 2.3 – 3.11 = - 0.81 (x14 – x\_mean)^2 = 0.65

i = 15 x15 – x\_mean = 3.0 – 3.11 = -0.11 (x15 – x\_mean)^2 = 0.01

sum((xi – x\_mean)^2) = 1.72

sigma^2 = sum((xi – x\_mean)^2) /15 = 0.11

sigma = 0.33

**z\_i = new x\_i = (xi-x\_mean)/sigma**

**sepal width:**

new x\_1 = (x1-x\_mean)/sigma = -0.21/0.33 = -0.64

new x\_2 = (x2 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_3 = (x3 - x\_mean)/sigma = - 0.61/0.33 = - 1.85

new x\_4 = (x4 - x\_mean)/sigma = 0.39/0.33 = 1.18

new x\_5 = (x5 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_6 = (x6 - x\_mean)/sigma = -0.21/0.33 = -0.64

new x\_7 = (x7 - x\_mean)/sigma = 0.19/0.33 = 0.58

new x\_8 = (x8 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_9 = (x9 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_10 = (x10 - x\_mean)/sigma = 0.09/0.33 = 0.27

new x\_11 = (x11 - x\_mean)/sigma = 0.19/0. 0.33 = 0.58

new x\_12 = (x12 - x\_mean)/sigma = -0.21/0. 0.33 = -0.64

new x\_13 = (x13 - x\_mean)/sigma = 0.09/0.33 = 0.27

new x\_14 = (x14 - x\_mean)/sigma = - 0.81/0.33 = -2.45

new x\_15 = (x15 - x\_mean)/sigma = -0.11/0.33 = -0.33

**(B)** Use the k-nearest neighbours method with **k**=3 to predict the missing classes. Use the Euclidean norm in your distance calculations.

new x\_1 = (x1-x\_mean)/sigma = -0.21/0.33 = -0.64

new x\_2 = (x2 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_3 = (x3 - x\_mean)/sigma = - 0.61/0.33 = - 1.85

new x\_4 = (x4 - x\_mean)/sigma = 0.39/0.33 = 1.18

new x\_5 = (x5 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_6 = (x6 - x\_mean)/sigma = -0.21/0.33 = -0.64

new x\_7 = (x7 - x\_mean)/sigma = 0.19/0.33 = 0.58

new x\_8 = (x8 - x\_mean)/sigma = 0.29/0.33 = 0.88

new x\_9 = (x9 - x\_mean)/sigma = 0.29/0.33 = 0.88

**Let's enter new values into the table (train data). New values are the mean of converted values.**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| 0.58 | -0.64 | virginica |
| -0.78 | 0.88 | setosa |
| - 0.10 | - 1.85 | virginica |
| - 0.9 | 1.18 | setosa |
| - 1.12 | 0.88 | setosa |
| 0.92 | -0.64 | virginica |
| 0.58 | 0.58 | virginica |
| 0.58 | 0.88 | virginica |
| 0.24 | 0.88 | virginica |

**Let's enter new values into the table (test data). New values are the mean of converted values.**

**We need to find a class.**

**k=3**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| -1.24 | 0.27 |  |
| -0.9 | 0.58 |  |
| 0.35 | -0.64 |  |
| 1.14 | 0.27 |  |
| -1.47 | -2.45 |  |
| 2.17 | -0.33 |  |

To calculate the distance, we use the formula:

**dist(i, j) = sqrt((x1i - x1j)^2 + (x2i - x2j)^2)**

**i - index of the element from train data**

**j - index of the element from test data**

**1 – value from the 1-st column**

**2 – value from the 2-nd column**

**Consider the first test element :**

|  |  |
| --- | --- |
| -1.24 | 0.27 |

**dist(1, 1)** =sqrt((-1.24-0.58)^2 + (0.27-(-0.64))^2)=sqrt(3.31 + 0.91) = 2.05

**dist(2, 1)** =sqrt((-1.24+0.78)^2 + (0.27-(0.88))^2)=0.76

**dist(3, 1)** =sqrt((-1.24+0.10)^2 + (0.27+1.85)^2)= 2.41

**dist(4, 1)** =sqrt((-1.24+0.9)^2 + (0.27-1.18)^2)= 0.97

**dist(5, 1)** =sqrt((-1.24+1.12)^2 + (0.27-0.88)^2)= 0.62

**dist(6, 1)** =sqrt((-1.24-0.92)^2 + (0.27+0.64)^2)= 2.34

**dist(7, 1)** =sqrt((-1.24-0.58)^2 + (0.27-0.58)^2)= 1.85

**dist(8, 1)** =sqrt((-1.24-0.58)^2 + (0.27-0.88)^2)= 1.92

**dist(9, 1)** =sqrt((-1.24-0.24)^2 + (0.27-0.88)^2)= 1.60

**Finding 3 minimum results:** **dist(5, 1)=**0.62**, dist(2, 1)=**0.76**, dist(4, 1)=**0.97

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(2, 1)** | setosa |
| **dist(4, 1)** | setosa |
| **dist(5, 1)** | setosa |

We can conclude that our first test instance belongs to the class **setosa**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| -1.24 | 0.27 | setosa |

**Consider the second test element :**

|  |  |
| --- | --- |
| -0.9 | 0.58 |

**dist(1, 2)** =sqrt((-0.9-0.58)^2 + (0.58-(-0.64))^2)=1.92

**dist(2, 2)** =sqrt((-0.9+0.78)^2 + (0.58-(0.88))^2)=0.32

**dist(3, 2)** =sqrt((-0.9+0.10)^2 + (0.58+1.85)^2)= 2.56

**dist(4, 2)** =sqrt((-0.9+0.9)^2 + (0.58-1.18)^2)= 0.6

**dist(5, 2)** =sqrt((-0.9+1.12)^2 + (0.58-0.88)^2)= 0.37

**dist(6, 2)** =sqrt((-0.9-0.92)^2 + (0.58+0.64)^2)= 2.19

**dist(7, 2)** =sqrt((-0.9-0.58)^2 + (0.58-0.58)^2)= 1.48

**dist(8, 2)** =sqrt((-0.9-0.58)^2 + (0.58-0.88)^2)= 1.51

**dist(9, 2)** =sqrt((-0.9-0.24)^2 + (0.58-0.88)^2)= 1.18

**Finding 3 minimum results:** **dist(2, 2**)= 0.32**, dist(4, 2)** = 0.6**, dist(5, 2)** =0.37

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(2, 2)** | setosa |
| **dist(4, 2)** | setosa |
| **dist(5, 2)** | setosa |

We can conclude that our second test instance belongs to the class **setosa**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| -0.9 | 0.58 | setosa |

**Consider the third test element :**

|  |  |
| --- | --- |
| 0.35 | -0.64 |

**dist(1, 3)** =sqrt((0.35-0.58)^2 + (-0.64-(-0.64))^2)=0.23

**dist(2, 3)** =sqrt((0.35+0.78)^2 + (-0.64-(0.88))^2)=1.89

**dist(3, 3)** =sqrt((0.35+0.10)^2 + (-0.64+1.85)^2)= 1.29

**dist(4, 3)** =sqrt((0.35+0.9)^2 + (-0.64-1.18)^2)= 2.20

**dist(5, 3)** =sqrt((0.35+1.12)^2 + (-0.64-0.88)^2)= 2.11

**dist(6, 3)** =sqrt((0.35-0.92)^2 + (-0.64+0.64)^2)= 0.57

**dist(7, 3)** =sqrt((0.35-0.58)^2 + (-0.64-0.58)^2)= 1.24

**dist(8, 3)** =sqrt((0.35-0.58)^2 + (-0.64-0.88)^2)= 1.54

**dist(9, 3)** =sqrt((0.35-0.24)^2 + (-0.64-0.88)^2)= 1.52

**Finding 3 minimum results: dist(1, 3)** = 0.23**, dist(6, 3)** = 0.57**, dist(7, 3)** = 1.24

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(1, 3)** | virginica |
| **dist(6, 3)** | virginica |
| **dist(7, 3)** | virginica |

We can conclude that our third test instance belongs to the class **virginica**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| 0.35 | - 0.64 | **virginica** |

**Consider the fourth test element :**

|  |  |
| --- | --- |
| 1.14 | -0.27 |

**dist(1, 4)** =sqrt((1.14-0.58)^2 + (0.27-(-0.64))^2)=1.07

**dist(2, 4)** =sqrt((1.14+0.78)^2 + (0.27-(0.88))^2)= 2.01

**dist(3, 4)** =sqrt((1.14+0.10)^2 + (0.27+1.85)^2)= 2.46

**dist(4, 4)** =sqrt((1.14+0.9)^2 + (0.27-1.18)^2)= 2.23

**dist(5, 4)** =sqrt((1.14+1.12)^2 + (0.27-0.88)^2)= 2.34

**dist(6, 4)** =sqrt((1.14-0.92)^2 + (0.27+0.64)^2)= 0.94

**dist(7, 4)** =sqrt((1.14-0.58)^2 + (0.27-0.58)^2)= 0.64

**dist(8, 4)** =sqrt((1.14-0.58)^2 + (0.27-0.88)^2)= 0.83

**dist(9, 4)** =sqrt((1.14-0.24)^2 + (0.27-0.88)^2)= 1.09

**Finding 3 minimum results: dist(6, 4)** = 0.94**, dist(7, 4)** = 0.64**, dist(8, 4)** = 0.83

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(6, 4)** | virginica |
| **dist(7, 4)** | virginica |
| **dist(8, 4)** | virginica |

We can conclude that our fourth test instance belongs to the class **virginica**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| 1.14 | 0.27 | **virginica** |

**Consider the fifth test element :**

|  |  |
| --- | --- |
| -1.47 | -2.45 |

**dist(1, 5)** =sqrt((-1.47-0.58)^2 + (-2.45-(-0.64))^2)= 2.73

**dist(2, 5)** =sqrt((-1.47+0.78)^2 + (-2.45 -(0.88))^2)= 3.4

**dist(3, 5)** =sqrt((-1.47+0.10)^2 + (-2.45+1.85)^2)= 1.50

**dist(4, 5)** =sqrt((-1.47+0.9)^2 + (-2.45-1.18)^2)= 3.67

**dist(5, 5)** =sqrt((-1.47+1.12)^2 + (-2.45-0.88)^2)= 3.35

**dist(6, 5)** =sqrt((-1.47-0.92)^2 + (-2.45+0.64)^2)= 2.99

**dist(7, 5)** =sqrt((-1.47-0.58)^2 + (-2.45-0.58)^2)= 3.66

**dist(8, 5)** =sqrt((-1.47-0.58)^2 + (-2.45-0.88)^2)= 3.91

**dist(9, 5)** =sqrt((-1.47-0.24)^2 + (-2.45-0.88)^2)= 3.74

**Finding 3 minimum results:**

**dist(3, 5)** = 1.5**, dist(1, 5)** = 2.73**, dist(6, 5)** = 2.99

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(1, 5)** | virginica |
| **dist(3, 5)** | virginica |
| **dist(6, 5)** | virginica |

We can conclude that our fifth test instance belongs to the class **virginica**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| -1.47 | -2.45 | **virginica** |

**Consider the sixth test element :**

|  |  |
| --- | --- |
| 2.17 | -0.33 |

**dist(1, 6)** =sqrt((2.17-0.58)^2 + (-0.33-(-0.64))^2)= 1.62

**dist(2, 6)** =sqrt((2.17+0.78)^2 + (-0.33-(0.88))^2)= 3.19

**dist(3, 6)** =sqrt((2.17+0.10)^2 + (-0.33+1.85)^2)= 2.73

**dist(4, 6)** =sqrt((2.17+0.9)^2 + (-0.33-1.18)^2)= 3.42

**dist(5, 6)** =sqrt((2.17+1.12)^2 + (-0.33-0.88)^2)= 3.50

**dist(6, 6)** =sqrt((2.17-0.92)^2 + (-0.33+0.64)^2)= 1.29

**dist(7, 6)** =sqrt((2.17-0.58)^2 + (-0.33-0.58)^2)= 1.83

**dist(8, 6)** =sqrt((2.17-0.58)^2 + (-0.33-0.88)^2)= 1.99

**dist(9, 6)** =sqrt((2.17-0.24)^2 + (-0.33-0.88)^2)= 2.28

**Finding 3 minimum results:**

**dist(1, 6)** = 1.62, **dist(6, 6)** = 1.29, **dist(7, 6)** = 1.83

We know what class they belong to:

|  |  |
| --- | --- |
| **dist(1, 6)** | virginica |
| **dist(6, 6)** | virginica |
| **dist(7, 6)** | virginica |

We can conclude that our sixth test instance belongs to the class **virginica**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| 2.17 | -0.33 | **virginica** |

**Results:**

|  |  |  |
| --- | --- | --- |
| **sepal length** | **sepal width** | **class** |
| -1.24 | 0.27 | **setosa** |
| -0.9 | 0.58 | **setosa** |
| 0.35 | -0.64 | **virginica** |
| 1.14 | 0.27 | **virginica** |
| -1.47 | -2.45 | **virginica** |
| 2.17 | -0.33 | **virginica** |