1 Exercise *(k-nearest Neighbor (4p))*

Mrs.A who studies Cognitive Science is looking for a T-shirt for her boyfriend, whose weight is about 80 kg and 177 cm tall. Please help her to find the right T-shirt size using simple k-Nearest Neighbor and Euclidean distance. To be certain, pick k=1,3 and 5.

Distances of $x = (177, 80)$ to each other data point:

$$
\begin{aligned}
d(x, x_1) &= \sqrt{(177-188)^2 + (80-100)^2} &&= \sqrt{521} = 22.8254 \\
d(x, x_2) &= \sqrt{(177-178)^2 + (80-108)^2} &&= \sqrt{785} = 28.0178 \\
d(x, x_3) &= \sqrt{(177-170)^2 + (80-50)^2} &&= \sqrt{949} = 30.8058 \\
d(x, x_4) &= \sqrt{(177-180)^2 + (80-86)^2} &&= 3\sqrt{5} = 6.7082 \\
d(x, x_5) &= \sqrt{(177-193)^2 + (80-70)^2} &&= 2\sqrt{89} = 18.868 \\
d(x, x_6) &= \sqrt{(177-182)^2 + (80-61)^2} &&= \sqrt{386} = 19.6469 \\
d(x, x_7) &= \sqrt{(177-187)^2 + (80-70)^2} &&= 10\sqrt{2} = 14.1421 \\
d(x, x_8) &= \sqrt{(177-173)^2 + (80-93)^2} &&= \sqrt{185} = 13.6015 \\
d(x, x_9) &= \sqrt{(177-172)^2 + (80-80)^2} &&= 5 \\
d(x, x_{10}) &= \sqrt{(177-185)^2 + (80-92)^2} &&= 4\sqrt{13} = 14.4222 \\
d(x, x_{11}) &= \sqrt{(177-174)^2 + (80-80)^2} &&= 3 \\
d(x, x_{12}) &= \sqrt{(177-174)^2 + (80-70)^2} &&= \sqrt{109} = 10.4403
\end{aligned}
$$

Since we are dealing with discrete valued output, we take the target value that occurs most often among the k nearest neighbors as the target value for x.

- $k = 1$-nearest neighbors:
  $x_{11} = (174, 80), t_{11} = XL$
  Choose $t = XL$.

- $k = 3$-nearest neighbors:
  $x_{11} = (174, 80), t_{11} = XL$
  $x_9 = (172, 80), t_9 = XL$
  $x_4 = (180, 86), t_4 = M/L$
  Choose $t = XL$.

- $k = 5$-nearest neighbors:
  $x_{11} = (174, 80), t_{11} = XL$
  $x_9 = (172, 80), t_9 = XL$

$x_4 = (180, 86), t_4 = M/L$
$x_{12} = (174, 70), t_{12} = M/L$
$x_8 = (173, 93), t_8 = XL$
Choose $t = XL$.

2 Exercise *(RBF (8p))*

1. Discuss RBF network and MLP in different aspects e.g. input and output dimension, extrapolation, lesion tolerance and advantages of each network.

   While both MLP and RBF network take the number of example features as their input dimension, the output of a RBF network is one-dimensional while that of a MLP may be multi-dimensional.

   The RBF network is a local method, meaning that a single adaptation step will only have an effect on weights in a certain subregion of the input space. Hence, even if a neuron is not functional anymore, the RBF network is barely affected. In contrast, a single adaptation step in a MLP results in an update of *all* weights and thus, a single missing neuron might prevent the whole network from functioning correctly. In short, the RBF network is "tolerant against lesions" and the MLP is not.

   Additionally, the RBF network has the advantage that its parameters are in general easier to choose, to handle and to interpret than the parameters of the MLP: When it comes to architectural parameters in the RFB network, one only has to decide on a number of basis functions. In the MLP one must choose an appropriate number of layers and number of hidden neurons, both of which can greatly affect the performance of the network. Moreover, the effects of a change in adaptation parameters of a RBF network (clustering parameters, radii, stepsize) are easy to predict. Changing the parameters in a MLP such as stepsize or momentum, on the other side, may have unforeseen consequences.

2. The training of RBF network concerns three parts. The first step is to find suitable centers or input weights, ξ. Explain in detail how to find these input weights.

   A simple way of finding input weights is using the training examples themselves: $\xi_i = x_i$. Another possibility is to perform clustering on the given examples x_i and take, e.g., the resulting centroids as input weights. Alternatively, one can perform Expectation Maximization on all parameters (input weights ξ_i, radii σ_i and output weights w_i) at the same time.

3. Write down another basis function which has the property $\Phi(r) \to 0$ as $|r| \to \infty$ and one example a of basis function which has property: $\Phi(r) \to \infty$ as $|r| \to \infty$.

3 Exercise *(SOM (8p))*

1. Explain

   (a) the meaning of topology preservation:

   (b) the properties of the topology function:

   (c) measuring similarity in SOM:

2. How to avoid that the later training phases forcefully pull the entire map towards a new pattern?

3. Briefly discuss at least three applications of SOM in different aspects.

   Dimension reduction using principal curves. Clustering. Visualization.