Machine Learning
Exercise Sheet 6

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

November 30th, 2017
Submission until December 7th (noon) by dropping at MACHINE LEARNING postbox
(please indicate in which tutorial are you participating!)
The postboxes is located inside the Samelsonplatz building to the right.

Exercise 11: Distance Metrics (10 Points)

Given are 5 Locations with following coordinates x1 and x2:

No.	$City$	x_1	x_2
1	Kakariko Village	4	2
2	Zebes	100	80
3	Donut Plains	11	5
4	Corneria	6	4
5	Pallet Town	1	1

The distance between the city x with coordinates (x_1, x_2) and the city $\tilde{x}$ with coordinates $(\tilde{x}_1, \tilde{x}_2)$ is defined by:

$$d(x, \tilde{x}) = \sqrt{(x_1 - \tilde{x}_1)^2 + (x_2 - \tilde{x}_2)^2}$$

a) Estimate the distance matrix D for the 5 cities.

b) Some students are developing a fighting computer game called ”Machine Learning Break-Everything-In-Our-Way”. In order to balance the difficulty of playing with each character, they compared the button sequences between characters in order for them to perform special attacks. This comparison is done by calculating the Edit Distance to measure how similar they are. Execute the algorithm introduced in class to calculate the edit distance of the following combo sequences:

AXBXAABX
XBAXXBX

Exercise 12: Nearest-Neighbor and Kernel Regression (10 Points)

Given is following data set:

x	y
1	20
2	18
3	16
4	14
5	12
6	10
7	8
8	6
9	4
10	2

a) Predict the target for $x = 0, x = 2.5$ and $x = 5.75$ using 2-nearest-neighbor regression using the L_2 metric.

b) Make a sketch of the final prediction for $x \in [0, 10]$ of the resulting 2-nearest-neighbor regression. What is noticeable?

c) The nearest-neighbor regression considers instances in its neighborhood, but neglects the actual distance. Kernel regression is similar to nearest-neighbor regression where the neighborhood **does not have a fixed size**. Instead, all instances contribute to the final prediction weighted by their similarity to the instance for which we want to predict. Precisely, the prediction is

$$\hat{y}(x_0) = \frac{\sum_{(x_i,y_i)\in\mathcal{D}_{train}} K(x_i,x_0)y_i}{\sum_{(x_i,y_i)\in\mathcal{D}_{train}} K(x_i,x_0)}$$

where K is a similarity measure. So the prediction is an average of the targets seen in the training data, however weighted by the similarities. Use the similarity function

$$K(x, x_0) = D\left(\frac{|x-x_0|}{\lambda}\right)$$

where $D(t)$ is defined as

$$D(t) = \begin{cases} \frac{3}{4}(1-t^2) & t < 1 \\ 0 & otherwise \end{cases}$$

and $\lambda = 2$ to predict the target for $x = 0, x = 2.5$ and $x = 5.75$.