Deadline: Friday December 11th , 10:00 Upload a `.pdf` file via LearnWeb. (e.g. exported Jupyter notebook)

1 Distance Measures (10+3 points)

A [3p] Compute the **Levenshtein distance** between the two strings MACHINE and LEARNING

		M	A	C	H	I	N	E
	0	1	2	3	4	5	6	7
L	1							
E	2							
A	3							
R	4							
N	5							
I	6							
N	7							
G	8							

B [3p] Show that the **Hamming distance** between two sets

$$\text{dist}_{\text{Ham}}(X,Y) = |(X \setminus Y) \cup (Y \setminus X)| = |(X \cup Y) \setminus (Y \cap X)|$$

satisfies the 3 properties (positive definiteness, symmetry and triangle inequality) a of distance measures.

Hint: Draw some Venn diagrams!

C [4p] In $\mathbb{R}^2$, draw all points that are distance 1 away from the origin with respect to

1. The taxicab distance $\text{dist}(x,y) = \|x-y\|_1$
2. The euclidean distance $\text{dist}(x,y) = \|x-y\|_2$
3. The maximum distance $\text{dist}(x,y) = \|x-y\|_\infty$

4.* The Mahalanobis distance with matrix $\Sigma^{-1} = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$

[3] Given a symmetric, positive definite covariance matrix Σ, show that the **Mahalanobis distance**

$$\text{dist}_{\text{Maha}}(x,y) = \sqrt{(x-y)^\intercal \Sigma^{-1} (x-y)}$$

satisfies the 3 defining properties of a distance measure.

Hint: Try to make use of the following facts from linear algebra.

- If $\langle x|y \rangle$ is an **inner product**, then $\|x\| \stackrel{\text{def}}{=} \sqrt{\langle x|x \rangle}$ is a **norm**.
- If $\|x\|$ is a **norm**, then $\text{dist}(x,y) \stackrel{\text{def}}{=} \|x-y\|$ is a **distance measure**.

2 K Nearest Neighbors (10 points)

Given is following data set:

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

A [3p] Predict the target for $x = 0, x = 2.5$ and $x = 5.75$ using 2-nearest-neighbor regression using the L_2 metric.

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

C [5p] 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 $x = 5.75$.