Deadline: Th. Dec. 6, 10:00 am Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz, or upload them as `.pdf` or `.ipynb` files onto the LearnWeb.

Choose and solve 2 of the 3 exercises! You can work on all of them for bonus points. But you can only earn a maximum of 20+5 points on this sheet!

Exercise 1 (Nearest Neighbors - 10 points)**.** In this exercise we want to test K-Nearest-Neighbor (KNN) regression on the function

$$y = f(x) + \epsilon, \qquad f(x) = x, \qquad \epsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.1) \tag{1}$$

To setup, generate a training set $D_{\text{train}} = (X_{\text{train}}, Y_{\text{train}})$ with $N = 100$ data points, where x_i is randomly sampled from $[-1, +1]$. Use the taxicab distance.

1. (4) Plot the training set and the prediction of the KNN regression with $K = 10$ on $[-2, 2]$. What happens at the edges?
2. (2) What happens for small/large values of K?
3. (4) Create a test set $D_{\text{test}} = (X_{\text{test}}, Y_{\text{test}})$ with $N = 100$ data points, where x_i is randomly sampled from $[-1, +1]$. For $K = 1..50$ compute both the loss $\text{MSE}(\hat{y}(X_{\text{train}}), Y_{\text{train}})$ and the score $\text{MSE}(\hat{y}(X_{\text{test}}), Y_{\text{test}})$. Plot both of them. What is the optimal value for K?

Exercise 2 (Distance metrics - 10 points)**.**

1. (3) Draw all points that are distance 1 away from the origin in $\mathbb{R}^2$ according to (1) the taxicab distance, (2) the euclidean distance and (3) the maximum distance.
2. (3) Show that the taxicab distance satisfies the defining properties of being a metric.
3. (4) Compute the Levenshtein distance between `HAPPY` and `HIPPO` by filling out the table according to the recursive algorithm.

		H	A	P	P	Y
	0	1	2	3	4	5
H	1					
I	2					
P	3					
P	4					
O	5					

Exercise 3 (Distance Measures - 10 points)**.** In lecture 3 the *Mahalanobis distance*

$$\mathrm{d}(x,y) = \sqrt{(x-y)^\intercal \Sigma^{-1}(x-y)} \tag{2}$$

was introduced in the context of LDA. In this exercise we want to show that this is indeed a metric.

1. (3) Show that if A is spd,[1] then $\langle x|y\rangle_A \overset{\text{def}}{=} x^\intercal Ay$ is an inner product[2]
2. (4) Show that if $\langle\cdot|\cdot\rangle$ is an inner product, then $\|x\| \overset{\text{def}}{=} \sqrt{\langle x|x\rangle}$ is a norm.
3. (3) Show that if $\|\cdot\|$ is a norm, then $\mathrm{d}(x,y) \overset{\text{def}}{=} \|x-y\|$ is a metric.

It follows that for any spd matrix A

$$\mathrm{d}(x,y) = \sqrt{(x-y)^\intercal A(x-y)} \tag{3}$$

is a metric. The inner product $\langle x|y\rangle_A \overset{\text{def}}{=} x^T Ay$ has important applications beyond the Mahalanobis distance like for example the conjugate gradient method[3] in convex optimization.

Hint: Base your arguments on the defining properties of the inner product, norm and metric (see table). To prove the subadditivity in part 2, use the Cauchy-Schwartz-Inequality $|\langle x|y\rangle|^2 \leq \langle x|x\rangle \cdot \langle y|y\rangle$, which holds true for any inner product!

inner product $\langle\cdot|\cdot\rangle$	norm $\|\cdot\|$	metric $\mathrm{d}(\cdot,\cdot)$
$\langle \alpha x|y\rangle = \alpha\langle x|y\rangle$ $\langle x+z|y\rangle = \langle x|y\rangle + \langle z|y\rangle$ (linearity)	$\|\alpha x\| = |\alpha|\cdot\|x\|$ (absolute homogeneity)	–
$\langle x|y\rangle = \langle y|x\rangle$ (symmetry)	–	$\mathrm{d}(x,y) = \mathrm{d}(y,x)$ (symmetry)
–	$\|x+y\| \leq \|x\| + \|y\|$ (subadditivity)	$\mathrm{d}(x,z) \leq \mathrm{d}(x,y) + \mathrm{d}(y,z)$ (subadditivity)
$\langle x|x\rangle \geq 0$ $\langle x|x\rangle = 0$ iff $x=0$ (positive definiteness)	$\|x\| \geq 0$ $\|x\| = 0$ iff $x=0$ (positive definiteness)	$\mathrm{d}(x,y) \geq 0$ $\mathrm{d}(x,y) = 0$ iff $x=y$ (positive definiteness)

Table 1: defining properties of inner products, norms, and metrics

[1] symmetric **p**ositive **d**efinite, cf. tutorial 2
[2] Note that in the case $A = I$, this is simply the standard dot product!
[3] https://en.wikipedia.org/wiki/Conjugate_gradient_method

Note: The word "iff" is a short form for "if and only if". The given properties must hold for all vectors $x, y, z \in \mathbb{R}^n$ and scalars $\alpha \in \mathbb{R}$. Finally, note that due to symmetry the inner product is not only linear in the first component, but in both!