Deadline: Th. Mai 2, 10:00 am Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz. Alternatively upload a Jupyter notebook (`.ipynb`) or a `.pdf` file via LearnWeb.

1 Gaussian Processes (programming) (15 points)

A. [5p] Implement a GP model take can perform regression for 1-dimensional input data, (preferably in PYTHON) using the squared exponential kernel (1).

$$\kappa(x, x') = \sigma_f^2 e^{-\frac{1}{2\ell^2}\|x-x'\|^2} \tag{1}$$

It should be a class with at least 4 functions:

`init`$(\sigma_f, \sigma_y, \ell)$:	Initialize the model with the given parameters
`fit`(X, Y):	fit the model to the training data
	(compute $K_y = K + \sigma_y^2 I$ and $\alpha = K_y^{-1} Y$)
`predict`(X):	Compute the prediction (μ_*, Σ_*)
`evaluate`(X, Y):	Compute the empirical loss $L = \|Y - \widehat{Y}\|^2$

B. [5p] Perform a GP regression on the data-set `tutorial2.dat` provided via LearnWeb. Use the parameters $\sigma_f = 1, \sigma_y = 0.5, \ell = 1$. Plot the prediction, including a 2σ uncertainty margin (cf. lecture slide plots) on the interval $[-3, +3]$.

C. [5p] Implement another method:

`optimize`$(\eta, \text{maxiter}, \text{tol})$:	Computes the optimal parameters σ_f, σ_y, ℓ
	by maximizing the marginal likelihood via Gradient Ascent

Use this function to estimate the optimal parameters σ_f, σ_y, ℓ for the `tutorial2.dat` data set. Try to achieve a tolerance of $\|\nabla_\theta \mathcal{L}\| < \text{tol} = 10^{-4}$ for the gradient of the marginal likelihood. Compute the resulting empirical loss on the training set and plot the prediction. Compare against the ground truth $f(x) = 2\sin(2x)e^{-\frac{1}{2}x}$

2 Gaussian Processes II (5 points)

A. [3p] Consider a GP with the Gaussian Kernel

$$\kappa(x, x') = \sigma_f^2 e^{-\frac{1}{2\ell^2}\|x-x'\|^2} \tag{2}$$

Consider the dataset consisting of the single data-point $x = 0, y = 0$. Plot the posterior and a 95% ($= 2\sigma$) confidence interval (uncertainty margin) around it. By experimenting, find how the posterior is influenced by the parameters (σ_f, σ_y and l)

Provide a geometrical interpretation for the effect these parameters have on the uncertainty margin.

B. [2p] Calculate Σ_{**} in two cases: $x' = x$ and for x' "far away" from x (i.e. $|x - x'| \to \infty$). Does it match the results from part A?