MACHINE LEARNING FUNDAMENSTALS - LS2026
HOMEWORK: NESTED CROSS-VALIDATION

Assignment (4 points) *Assume you are given a set of i.i.d. generated examples*

$$T_n = \big((\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \{-1, +1\} \mid i = 1, \ldots, n\big) \ ,$$

Your goal is to learn a Support Vector Machine (SVM) classifier with the Radial Basis Function (RBF) kernel

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \left\|\mathbf{x} - \mathbf{x}'\right\|^2) \ .$$

The script should output the trained SVM classifier together with a 95% *confidence interval (CI) on its true classification error.*

The regularization constant C *is unknown and its optimal value should be tuned from the set* $\{0.1, 1.0, 10.0, 100.0\}$. *The RBF kernel width should be set to*

$$\gamma = \frac{1}{d \cdot \sigma^2} \ ,$$

where d *is the input vector dimension and* σ^2 *is the average variance of the input features.*

Both the regularization constant C *and the generalization performance of the resulting classifier should be estimated by nested cross-validation. Both the inner and output loop should use 5 folds. To compute the confidence interval, use the corrected resampled* t*-test of Nadeau & Bengio (2003).*

The solution should output:

1. *The RBF SVM classifier* $\hat{h}$ *with the tuned regularization constant* C.
2. *A confidence interval* $[L, U]$ *such that the true error of the trained classifier* $\hat{h}$ *lies in* $[L, U]$ *with probability* 95%.