Machine Learning Assignment 1
Linear Models & Kernel Method

Submission deadline: October 26, 2023.

Problem 1. Ridge Regression (15 points).

In a regression task, we have vectors $\boldsymbol{x} \in \mathbb{R}^D$, target values $y \in \mathbb{R}$ associated with them, and some model $f(\boldsymbol{x}) : \mathbb{R}^D \to \mathbb{R}$ to predict the target values for arbitrary vectors in $\mathbb{R}^D$.

Suppose we have a training dataset $\{\Phi, \boldsymbol{t}\}$, where $\Phi_{N\times D}$ is the design matrix in which each row is a feature vector $\phi(\boldsymbol{x})$ of a training point $\boldsymbol{x}$, $\boldsymbol{t}_{N\times 1}$ is the vector with target values for the training points. N is the number of points in the training dataset, D is the dimensionality of the feature space. Suppose that each entry in the last column of Φ is equal to 1.

Derive the closed form solution for the optimal parameters of a ridge regression model:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \phi(\boldsymbol{x})$$

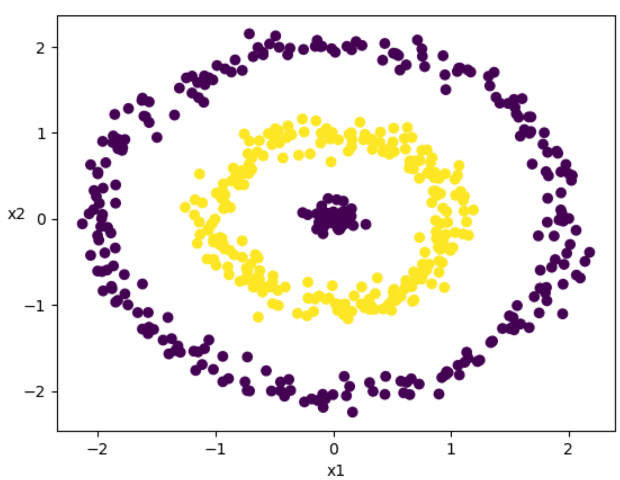
The optimal parameters give the minimum to the following loss function:

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}(f(\boldsymbol{x}_n) - t_n)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|^2$$

Here $\|\cdot\|$ is the Euclidean norm of a vector; $\phi(\boldsymbol{x}_n)$ and t_n are n-th rows of Φ and $\boldsymbol{t}$ respectively.

Problem 2. Feature engineering (10 points).

Suppose you have the following set S of 2D points, $S_n = (x_n^{(1)}, x_n^{(2)})$



Color denotes the class attribution of a point: blue points belong to the class C_1, yellow points belong to the class C_2.
Propose the new features for points in S based on $x^{(1)}$ and $x^{(2)}$. In this new feature space, classes C_1 and C_2 should be linearly separable.

Problem 3. Kernel functions (12 points).

Consider the following function $f : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$

$$f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{x} \boldsymbol{x}^T \boldsymbol{y} \boldsymbol{y}^T \boldsymbol{y}$$

Prove that f is a valid kernel or prove the opposite.

The only rules allowed to use without a proof:
$k : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ is a valid kernel if

$$k(\boldsymbol{x}, \boldsymbol{y}) = ck_1(\boldsymbol{x}, \boldsymbol{y})$$

$$k(\boldsymbol{x}, \boldsymbol{y}) = k_1(\boldsymbol{x}, \boldsymbol{y}) + k_2(\boldsymbol{x}, \boldsymbol{y})$$

$$k(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T A \boldsymbol{y}$$

$$k(\boldsymbol{x}, \boldsymbol{y}) = k_3(\phi(\boldsymbol{x}), \phi(\boldsymbol{y}))$$

where k_1 and k_2 are valid kernels in $\mathbb{R}^D$, $c > 0$ is a constant, ϕ is a function from $\mathbb{R}^D$ to $\mathbb{R}^M$, k_3 is a valid kernel in $\mathbb{R}^M$, A is a symmetric positive semidefinite matrix.

Problem 4. SVM (15 points).

Consider the following training data.

Class	x_1	x_2
$+$	1	1
$+$	2	2
$+$	2	0
$-$	1	-1
$-$	-1	0
$-$	0	1

1. Plot the six training points. Are the classes $\{+, -\}$ linearly separable?
2. Construct the weight vector of the maximum margin hyperplane by inspection and identify the support vectors.
3. If you remove one of the support vectors does the size of the optimal margin decrease, stay the same, or increase?
4. Is your answer to (3) also true for any dataset? Provide a counterexample or give a short proof.