Machine Learning Assignment 1
Linear Models & Kernel Methods

Submission deadline: October 26, 2024

1 Problem 1. Ridge Regression (10 points)

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

Suppose we have a training dataset $\{\mathbf{\Phi}, \mathbf{t}\}$, where $\mathbf{\Phi} \in \mathbb{R}^{N \times D}$ is the design matrix in which each row is a feature vector $\phi(\mathbf{x})$ of a training point $\mathbf{x}$, and $\mathbf{t} \in \mathbb{R}^{N \times 1}$ is the vector with target values for the training points. N is the number of points in the training dataset, and D is the dimensionality of the feature space. Suppose that each entry in the last column of $\mathbf{\Phi}$ is equal to 1. Your task is to derive the closed form solution for the optimal parameters of a ridge regression model.

- State the equation of a ridge regression model and identify the model parameters
- State the equation for the loss function (mean squared error) with an ℓ_2 regularization weighted by λ
- State which condition should be met in order to find the model parameters
- Find the ideal model parameters under the proposed loss function.

Note: You can use $\|\cdot\|$ as the Euclidean norm of a vector; $\phi(\mathbf{x}_n)$ and t_n are the n-th rows of $\mathbf{\Phi}$ and $\mathbf{t}$ respectively.

2 Problem 2. Feature Engineering (10 points) and Basic Concepts (10 points)

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

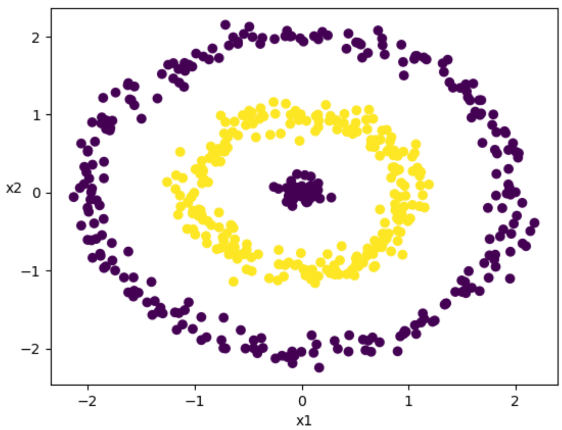

Figure 1: Color denotes the class attribution of a point: blue points belong to the class C_1, yellow points belong to the class C_2.

- Explain in detail 2 classification algorithms that could solve the problem. Discuss advantages and disadvantages of each.
- Propose 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. Come up with 2 different solutions meeting the stated criteria.

Problem 3. Kernel Functions (10 points)

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

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

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

The only rules allowed to use without proof are the following:

- Kernel functions are *linear* and positive.
- Kernel functions can be expressed as an inner product
- A kernel function of 2 inputs can be expressed as another kernel of a transformation of those inputs (into a potentially different space).

Begin by formalizing those rules and apply them to prove or disprove the statement.

Problem 4. SVM (10 points)

Consider the following training data:

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

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.