Exercises
Introduction to Machine Learning
SS 2019

Series 3, Mar 18th, 2019 (SVM, Kernels)

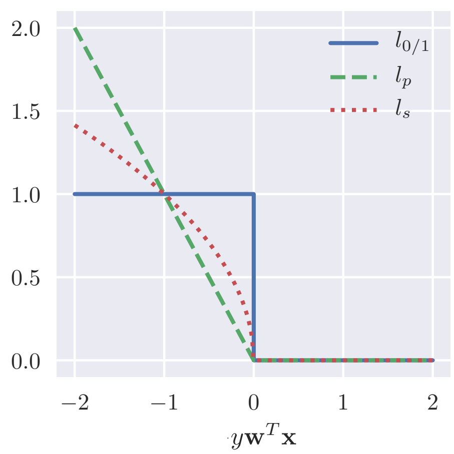
Institute for Machine Learning
Dept. of Computer Science, ETH Zürich
Prof. Dr. Andreas Krause
Web: `https://las.inf.ethz.ch/teaching/introml-s19`
Email questions to:
max.paulus@inf.ethz.ch, joanna.ficek@inf.ethz.ch

We will publish sample solutions on Monday, Mar 25th.

Problem 1 (SVM):

This exercise is based on an exercise designed by Stephanie Hyland. In its original formulation, the perceptron aims to minimise a $0/1$-loss function (shown below, solid). Because this objective is neither convex nor differentiable, a surrogate loss function is optimised (typically, $l_p(\mathbf{w}; \mathbf{x}, y) = \max(0, -y\mathbf{w}^T\mathbf{x})$, dashed). In this exercise, we consider a different surrogate loss function l_s, which approximates the $0/1$-loss function more closely.

$$l_s(\mathbf{w}; \mathbf{x}, y) = \begin{cases} 0, & \text{for } \mathrm{sign}(\mathbf{w}^T\mathbf{x}) = y \\ \sqrt{-y\mathbf{w}^T\mathbf{x}}, & \text{for } \mathrm{sign}(\mathbf{w}^T\mathbf{x}) \neq y \end{cases}$$



a) Is l_s convex? Is l_s differentiable? Justify your answer.

b) Derive $\nabla l_s(w, x, y)$

c) Write down pseudo-code for training an SVM using l_s.

Problem 2 (Kernels):

Using the basic rules for kernel decomposition discussed in class or otherwise and assuming that $k(x, y)$ is a valid kernel, letting $f : \mathbb{R} \to \mathbb{R}$ in a) and $f : \mathcal{X} \to \mathbb{R}$ for c) and d), and $\phi : \mathcal{X} \to \mathcal{X}'$, show that

a) $k_a(x, y) = f(k(x, y))$ is a valid kernel, if f is a polynomial with non-negative coefficients.

b) $k_b(x, y) = \exp(k(x, y))$ is a valid kernel.

c) $k_c(x, y) = f(x)k(x, y)f(y)$ is a valid kernel.

d) $k_d(x, y) = k(\phi(x), \phi(y))$ is a valid kernel.

Problem 3 (Past Exam):

1. For $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$, and $K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + 1)^2$, find a feature map $\phi(\boldsymbol{x})$, such that $k(\boldsymbol{x}, \boldsymbol{x'}) = \phi(\boldsymbol{x})^\top \phi(\boldsymbol{x}')$.
2. For the dataset $X = \{\boldsymbol{x}_i\}_{i=1,2} = \{(-3, 4), (1, 0)\}$ and the feature map $\phi(\boldsymbol{x}) = [x^{(1)}, x^{(2)}, \|\boldsymbol{x}\|]$, calculate the **Gram matrix** (for a vector $\boldsymbol{x} \in \mathbb{R}^2$ we denote by $x^{(1)}, x^{(2)}$ its components).