Exercises
Advanced Machine Learning
Fall 2019

Series 5, Nov 11th, 2019 (Support Vector Machines)

Machine Learning Laboratory
Dept. of Computer Science, ETH Zürich
Prof. Joachim M. Buhmann
Web https://ml2.inf.ethz.ch/courses/aml/

Teaching assistant: **Piazza (Francesco Locatello)**
locatelf@ethz.ch

Problem 1 (Warm-up: Kernel Function):

Say we are given the probability distribution function $p(\mathbf{x}, h)$, where $\mathbf{x} \in \mathcal{X}$ and $h \in \mathcal{H}$ (finite set). Consider the kernel $k\big((\mathbf{x}, h), (\mathbf{x}', h')\big)$ defined on pairs $(\mathbf{x}, h) \in \mathcal{X} \times \mathcal{H}$. Prove that $k_m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, with

$$k_m(\mathbf{x}, \mathbf{y}) = \sum_{h,h'} k\big((\mathbf{x}, h), (\mathbf{y}, h')\big) p(h \mid \mathbf{x}) p(h' \mid \mathbf{y}), \tag{1}$$

is a kernel function.

Problem 2 (SVMs as Nearest Neighbor Classifiers):

SVMs use the predictor function $f(\mathbf{x}) = \text{sign}\big(\sum_{i=1}^{n} \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i)\big)$, where $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ denotes the training set and $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a kernel function. Assume that $y_i \in \{-1, +1\}$ for $i = 1, \ldots, n$. Say $\alpha_i = 1$ for all $i = 1, \ldots, n$ and let $k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{h^2}\right)$, i.e. the RBF kernel.

Show that for all $\mathbf{x} \in \mathbb{R}^d$ which have a *unique nearest neighbour* amongst the points in $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, there exists an $h_0 > 0$ such that for all $h < h_0$ the resulting SVM prediction is the same as the prediction made by a Nearest Neighbor (1-NN) classifier.

Problem 3 (Dual Formulation for Structural SVM):

The primal formulation of soft-margin structural SVMs is as follows:

$$\begin{aligned} &\min_{\mathbf{w}, \xi \geq 0} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^{n} \xi_i \\ &\text{s.t. } \mathbf{w}^\top \Psi(z_i, \mathbf{y}_i) - \mathbf{w}^\top \Psi(z, \mathbf{y}_i) \geq \Delta(z, z_i) - \xi_i \quad \forall i, \ \forall z \neq z_i. \end{aligned} \tag{2}$$

To simplify the notation, let $\Psi_i(z) := \Psi(z_i, \mathbf{y}_i) - \Psi(z, \mathbf{y}_i)$ and $\Delta_i(z) := \Delta(z, z_i)$. Use the Lagrange method to derive the dual formulation of this optimization problem.