University of Salzburg — Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **D**

Exercise 1. 3 P.

Consider the domain $\mathcal{X} = \mathbb{R}^d$ and label set $\mathcal{Y} = \{-1, +1\}$. A 1-NN (1-nearest-neighbor) classifier assigns to a data point $\mathbf{x} \in \mathbb{R}^d$ the label of its closest (in Euclidean norm $\|\cdot\|$) training instance (i.e., a point from the training set S). Formally, given $S = ((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_m))$ training instances and a data point $\mathbf{x}$, we let $\pi_1(\mathbf{x}), \ldots, \pi_n(\mathbf{x})$ be a reordering of $\{1, \ldots, m\}$ such that

$$\forall i < m : \|\mathbf{x} - \mathbf{x}_{\pi_i(\mathbf{x})}\| \leq \|\mathbf{x} - \mathbf{x}_{\pi_{i+1}(\mathbf{x})}\| \ .$$

A 1-NN hypothesis, $h_S : \mathcal{X} \to \mathcal{Y}$, outputs

$$h_S(\mathbf{x}) = y_{\pi_1(\mathbf{x})} \ .$$

What is the VC dimension of the class of 1-NN classifiers (provide an argument, not just a solution).

Exercise 2. 5 P.

Let our domain be $\mathcal{X} = [0, 2\pi]$ and label set $\mathcal{Y} = \{-1, +1\}$. Consider the hypothesis class

$$\mathcal{H}_{\text{sin}} = \{f : \mathbf{X} \to \mathbf{Y}, x \mapsto f(x) = \text{sign}(\sin(wx)), w \geq 0\} \ .$$

Show that this class can assign the correct label (-1) for any set of negatively labeled (i.e., -1) points in

$$C_n = \{2\pi 10^i\}_{i=1}^n$$

for any $n > 0$ (similar arguments would hold for the positively labeled points). What do you conclude?

Exercise 3. 4 P.

Let our domain be $\mathcal{X} = \mathbb{R}^2$ and label set $\mathcal{Y} = \{-1, +1\}$. Consider the hypothesis class of axis-aligned rectangles

$$\mathcal{H}_{\text{rect}} = \{h_{l,r,t,b} : l < r, \text{and } b < t\}$$

(where l, r, t, b denotes left, right, top and bottom) with

$$h_{l,r,t,b}(\mathbf{x}) = \begin{cases} +1 & \text{if } l \leq x_1 \leq r \text{ and } b \leq x_2 \leq t \\ -1 & \text{otherwise} \end{cases}$$

(1) Find a set of four points that is shattered by this class (just draw the points and the corresponding rectangles) and (2) provide an argument that no set of five points is shattered by this class (does not have to be fully formal).

Total #points: 12 P.