University of Salzburg — Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **E**

Exercise 1. 4 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 $S = ((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n))$ training instances and a *new* data point $\mathbf{x}$, we let $\pi_1(\mathbf{x}), \dots, \pi_n(\mathbf{x})$ be a reordering of $\{1, \dots, n\}$ (according to their distance to $\mathbf{x}$) such that

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

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

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

What is the VC dimension of the class of 1-NN classifiers (provide an argument, not just a solution). Finally, does it even make sense to talk about the VC dimension for the 1-NN rule? (yes/no; explain your answer!)

Exercise 2. 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

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

(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).

Bonus! (related to nearest neighbor classification)

Exercise 3. 4 P.

Say you have a collection of subsets $A_1, \dots, A_k$ of some domain $\mathcal{X}$. Also, you have $x_1, \dots, x_m$ drawn i.i.d. from some distribution $\mathcal{D}$ over $\mathcal{X}$ and lets call this set of m points S. Show that

$$\mathbb{E}_{S \sim \mathcal{D}^m} \left[\sum_{i : A_i \cap S = \emptyset} \mathbb{P}[A_i] \right] \leq \frac{k}{me}$$

First, rewrite the sum as a sum over indicator functions; then use *linearity of expectation*; next, bound the sum $\sum_{i=1}^k \text{term}_i$ over the terms term_i by $k \max_i \text{term}_i$ and finally use the inequality $\max_x x e^{-mx} \leq 1/(me)$ (for $m > 0$) at the right place. You get another 2 points if you can also prove $\max_x x e^{-mx} \leq 1/(me)$.

Total #points: 12 P.