University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **A**

Background (Probability, inequalities, ...)

Exercise 1. 1 P.

Let X be a random variable that captures rolling a fair dice with 6 sides. Use Markov's inequality to bound $\mathbb{P}[X \geq 4]$ and also compute the exact probability. Is the bound loose or tight?

Exercise 2. 2 P.

Assume we have a random variable X which takes on values > -90 and we know $\mathbb{E}[X] = -30$. Bound $\mathbb{P}[X \geq -20]$.

Hint: How about defining a new (appropriate) random variable Y such that we can apply Markov's inequality?

Exercise 3. 2 P.

Consider the following problem with (binary) inputs $x_i \in \{0, 1\}$, $i = 1, \ldots, 4$ and (binary) output $y \in \{0, 1\}$:

x_1	x_2	x_3	x_4	y
1	0	0	1	0
1	1	0	1	1
1	0	1	1	1
1	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Lets say our learning objective is to learn a function

$$y = f(x_1, x_2, x_3, x_4)$$

which maps our four boolean inputs to one boolean output $y \in \{0, 1\}$. To get a feeling for the (size of the) problem, we want to know how many such functions exist? When you have the solution, think about what happens for n inputs $x_1, \ldots, x_n$? What do you think is the big problem here? Plot the number of functions as a function of the number of inputs n.

PAC Learning

Exercise 4. 2 P.

Given real numbers $a_1 \leq b_1$ and $a_2 \leq b_2$, define the predictor

$$h_{a_1,b_1,a_2,b_2}(x_1, x_2) = \begin{cases} 1 & \text{if } a_1 \leq x_1 \leq b_1 \text{ and } a_2 \leq x_2 \leq b_2 \\ 0 & \text{else} \end{cases}$$

This defines a rectangle in $\mathbb{R}^2$ which labels all points as 1 if they are inside and 0 otherwise. Assume realizability and let A be an algorithm that returns the smallest rectangle which encloses all positive instances in the training set S. *Argue that A is an ERM algorithm.*

Exercise 5. 3 P.

Let $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \{0, 1\}$ and consider hypotheses $h_r : \mathcal{X} \to \mathcal{Y}$ in $\mathcal{H}$ of the form

$$h_r(\boldsymbol{x}) = 1_{\|\boldsymbol{x}\| \leq r}(\boldsymbol{x}), \text{ with } r \in \mathbb{R}_+ \quad .$$

In other words, our hypotheses are *concentric circles*. Show that this class is PAC-learnable (i.e., assume realizability) from training data of size

$$m \geq \left(\frac{1}{\epsilon}\right) \log\left(\frac{1}{\delta}\right) \quad .$$

Total #points: 10 P.