University of Salzburg/Dept. of Computer Science | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **A**

Background (Probability, inequalities, ...)

Exercise 1. 2 P.

Let X be a random variable that captures the sum of rolling a fair dice 10 times. Use Chebychev's inequality to bound

$$\mathbb{P}[|X - \mu| \geq 10]$$

Hint: Note that all dice rolls are independent. Compute the expectation of $|X|$ to obtain μ, then the variance and go on from there (this is really just a calculus exercise)!

Exercise 2. 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 3. 3 P.

Let $\mathcal{H}$ be a hypothesis class of binary classifiers. Show that (1) if $\mathcal{H}$ is *agnostic PAC learnable*, then $\mathcal{H}$ is *PAC learnable* and (2) if A is a successful agnostic PAC learner for $\mathcal{H}$, then A is also a successful PAC learner for $\mathcal{H}$.

Hint: In agnostic PAC learning, $\mathcal{D}$ is a distribution over $\mathcal{X} \times \mathcal{Y}$. In PAC learning (i.e., we assume realizability), $\mathcal{D}$ is what? Now, squeeze this into the agnostic PAC learning framework.

Exercise 4. 5 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: 12 P.