University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **A**

Background (Probability, inequalities, ...)

Exercise 1. 5 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. 5 P.

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

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. To get a feeling for 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?

PAC Learning

Exercise 3. 5 P.

Let $\mathcal{H}$ be a class of binary classifiers over a domain $\mathcal{X}$, i.e., functions $h : \mathcal{X} \to \{0, 1\}$. Let $\mathcal{D}$ be an unknown distribution over $\mathcal{X}$ and let f be the target hypothesis in $\mathcal{H}$ (i.e., the true labeling function). Fix some $h \in \mathcal{H}$. Show that the expected value of $L_S(h)$ over the choice of $S|_x = (x_1, \ldots, x_m)$, with $|S| = m$, equals $L_{(\mathcal{D},f)}(h)$. In other words, show that

$$\mathbb{E}_{S|_x \sim \mathcal{D}^m}[L_S(h)] = L_{(\mathcal{D},f)}(h) \enspace .$$

Hint: Use the linearity of the expectation and the fact that the samples x_i are i.i.d. Start with:

$$\mathbb{E}_{S|_x \sim \mathcal{D}^m}[L_S(h)] = \mathbb{E}_{S|_x \sim \mathcal{D}^m}\left[\frac{1}{m}\sum_{i=1}^{m} 1_{h(x_i) \neq f(x_i)}\right]$$

Exercise 4. 10 P.

Let $\mathcal{H}$ be a hypothesis class of binary classifiers. Show that if $\mathcal{H}$ is *agnostic PAC learnable*, then $\mathcal{H}$ is *PAC learnable* as well. Furthermore, 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 (under realizability), $\mathcal{D}$ is what? Now, squeeze this into the agnostic PAC learning framework.

Exercise 5. 10 P.

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

$$h_r(\boldsymbol{x}) = 1_{\|\boldsymbol{x}\| \leq r}(\boldsymbol{x})$$

for some $r \in \mathbb{R}_+$ (i.e., our hypotheses are *concentric circles*). Show that this class is PAC-learnable from training data of size

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

under the assumption of *realizibility*.

Total #points: 35 P.