University of Salzburg | Lecturer: Roland Kwitt

Machine Learning (911.236)

Exercise sheet **C**

Exercise 1. 5 P.

Let $\mathcal{X}$ be our domain and $\mathcal{Y} = \{0, 1\}$ our label set. Further, let $D_1, \ldots, D_m$ be a sequence of distributions over $\mathcal{X}$. We define

$$\bar{D}_m = \frac{1}{m}(D_1 + \cdots + D_m) \ ,$$

and assume that our *finite* hypothesis class $\mathcal{H}$ of binary classifiers contains the true labeling function $f : \mathcal{X} \to \mathcal{Y}$. In other words $f \in \mathcal{H}$. Now, we are given a training set S of size m where the instances x_i are not identically distributed, but independent. Specifically, x_i is drawn from D_i and labeled by f, x_2 is drawn from D_2 and labeled by f, etc. Formally, we have

$$S = ((x_1, y_1), \ldots, (x_m, y_m)) \quad \text{with} \quad x_i \sim D_i \quad \text{and} \quad y_i = f(x_i) \ .$$

Now, fix $\epsilon \in (0, 1)$ and show

$$\mathbb{P}\left[\exists h \in \mathcal{H} \text{ s.t. } L_{\bar{D}_m, f}(h) > \epsilon \text{ and } L_S(h) = 0\right] \leq |\mathcal{H}| e^{-\epsilon m}$$

Strategy: Fix a *bad* hypothesis, i.e., one that has $L_{\bar{D}_m, f}(h) > \epsilon$, i.e.,

$$\frac{\mathbb{P}_{x \sim D_1}[h(x) \neq f(x)] + \cdots \mathbb{P}_{x \sim D_m}[h(x) \neq f(x)]}{m} > \epsilon \ .$$

We can now try to bound the probability that such a hypothesis achieves 0 empirical error on S, i.e., $L_S(h) = 0$. Along this way, we can use the *inequality of arithmetic and geometric means (AM-GM)*, i.e., for any nonnegative real numbers $x_1, \ldots, x_n$, it holds that

$$\frac{1}{n}(x_1 + \cdots + x_n) \geq (x_1 \cdots x_n)^{1/n} \ .$$

Total #points: 5 P.