COLT— Exercise Sheet 2

Exercise 1. (Singleton concepts) Let X be a discrete domain (countable), and let $\mathcal{C}_{sing} = \{h_z \mid z \in X\} \cup \{h_\perp\}$, where $h_z(x) = 1$ if and only if $x = z$, and $h_\perp(x) = 0$ for all $x \in X$. Assume that we have an algorithm for the sample fitting problem for $\mathcal{C}$. Show that $\mathcal{C}$ is efficiently PAC-learnable.

Exercise 2. (Conjunctions with few literals I) Present with details the algorithm for PAC-learning conjunctions with few literals in the following way, which has been sketched at the lecture.

Exercise 3. (Conjunctions with few literals II) Improve the algorithm for PAC-learning conjunctions with few literals in the following way:

The greedy (approximation) algorithm for the set cover problem finds a set cover of the size $l = opt(\mathcal{F}) \cdot m$ by taking l rounds. We can improve its efficiency, if we settle for approximation. Namely, assume that we execute the greedy algorithm as long as the uncovered part is at most $\frac{\epsilon}{2}m$. It follows that the algorithm returns *almost* set cover of the size at most $opt(\mathcal{F}) \cdot \log(\frac{2}{\epsilon})$.

Use the above observation to give an improved PAC-learning algorithm for conjunctions with few literals. Note that the concept returned by the algorithm may not be consistent with the sample. Show that it's generalization error is less than ϵ with the probability $1 - \delta$.

Exercise 4. Recall the example of learning axis-aligned rectangles in $\mathbb{R}^2$. The learning algorithm gets points $(x, y) \in R^2$ labeled positively (belong to the rectangle) or negatively. Now, consider the case, where positive examples are subjected to *noise*. More precisely, for each example (x, y), if (x, y) is outside of the rectangle we label it negatively. However, if it is inside, its label is positive with the probability $1 - p$ and negative with the probability p. The label of each positive example is probabilistically independent from all other labels.

Construct an efficient PAC-learning for learning rectangles in the presence of noise. Assume that the algorithm gets in input the upper bound $\mu < 0.5$ on the probability of noise, i.e., $p < \mu < 0.5$.