Problem 1

The submission deadline is 6pm on Thursday 1st March 2018. Late assignments will receive **no mark**. Your work should be handed in to me directly in class. You can choose to either type your homework, or write it by hand. Achtung! You assignment must be presented in an A4 paper, and no smaller. The form and neatness of work **will** be considered in marking. The total number of marks is 10. **Up to 5 marks out of a total of 10 can be (will be) removed from your final mark for lack of neatness**. Achtung! If the question says show.. then show.. if it says deduce, then deduce, if it says to recall a famous theorem, then give it. Working and/or reasoning **must** be given to obtain full credit. **Do not forget** to print your name on the first page.

PAC learnability.
In this problem, we establish that classes of functions $\mathcal{F}$ containing finitely many elements are PAC learnable. We recall that a class of functions $\mathcal{F}$ is called PAC learnable if there exists

(a) a function $n_{\mathcal{F}} : (0,1)^2 \to \mathbb{N}$, for which $n_{\mathcal{F}}(\epsilon, \delta)$, $\epsilon, \delta \in (0,1)$, is at most polynomial in ϵ^{-1} and δ^{-1} (for example, $n_{\mathcal{F}}(\epsilon, \delta) = \epsilon^{-2} \log(\delta^{-1})$ is such a function), and

(b) a predictor $\hat{f}_n \in \mathcal{F}$

such that for any $(\epsilon, \delta) \in (0,1)^2$, for all $n \geq n_{\mathcal{F}}(\epsilon, \delta)$, for any distribution of (X, Y), the inequality

$$\mathcal{R}(\hat{f}_n) \leq \inf_{f \in \mathcal{F}} \mathcal{R}(f) + \epsilon \,,$$

holds with probability larger than $1 - \delta$.

Consider a learning sample $\mathcal{L}_n := \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where each $(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ is independent and identically distributed (i.i.d.) and distributed like a generic $(X, Y) \sim \mathbf{P}$. Let $\mathcal{F}$ be a class of candidate functions $f : \mathbb{R}^d \to \mathbb{R}$, which predicts Y using $f(X)$. The risk of a fixed prediction function is $\mathcal{R}(f) := \mathbf{E}\left\{\ell(Y, f(X))\right\}$, for a loss function $\ell : \mathbb{R} \times \mathbb{R} \to [0,1]$. We consider minimisation of the empirical risk

$$\widehat{\mathcal{R}}_n(f) := \frac{1}{n} \sum_{i=1}^{n} \ell(Y_i,\, f(X_i)) \,,$$

leading to the definition of the empirical risk minimizer

$$\hat{f}_n := \arg\min_{f \in \mathcal{F}} \widehat{\mathcal{R}}_n(f) \,.$$

Finally, put $\bar{f} := \arg\min_{f \in \mathcal{F}} \mathcal{R}(f)$.

(i) Show that [3]

$$\mathcal{R}(\hat{f}_n) - \mathcal{R}(\bar{f}) \leq 2 \sup_{f \in \mathcal{F}} \left|\mathcal{R}(f) - \widehat{\mathcal{R}}_n(f)\right| .$$

Hint: Note that $\mathcal{R}(\hat{f}_n) = \mathcal{R}(\hat{f}_n) + \left\{\widehat{\mathcal{R}}_n(\hat{f}_n) - \widehat{\mathcal{R}}_n(\hat{f}_n)\right\} + \left\{\mathcal{R}(\bar{f}) - \mathcal{R}(\bar{f})\right\}$. This may be a good place to start.

(ii) Show using Hoeffding's inequality (that you will recall) that for any $\epsilon > 0$, [3]

$$\mathbf{P}\left(|\widehat{\mathcal{R}}_n(f) - \mathcal{R}(f)| \geq \epsilon\right) \leq 2e^{-2n\epsilon^2}\,.$$

(iii) Deduce from (ii) that if $\mathcal{F}$ contains finitely many elements, with cardinal $|\mathcal{F}|$, then for any $\epsilon > 0$, [1]

$$\mathbf{P}\left(\max_{f\in\mathcal{F}} |\widehat{\mathcal{R}}_n(f) - \mathcal{R}(f)| \geq \epsilon\right) \leq 2|\mathcal{F}|e^{-2n\epsilon^2}\,.$$

(iv) Conclude from (i) and (iii) that a finite class of function $\mathcal{F}$ is PAC learnable. (The definition of PAC learnability is also given on page 27 of the lecture notes SL: FOUNDATIONS.)

[3]

Total marks = 10