The following homework will be discussed on June 1st to 3rd:

I. AdaBoost

1. Consider a class of base hypotheses $\mathcal{F}$ with VC-dimension d and let $\mathcal{F}_T$ be the set of hypotheses that can be reached from $\mathcal{F}$ by AdaBoost after T rounds. Provide an upper bound on the growth function of $\mathcal{F}_T$ and of its VC-dimension in terms of d and T. (hint: for bounding the growth function regard the functions in $\mathcal{F}_T$ as a composition of two types of functions)

2. Let $p_i^{(t+1)}$, $t \in \{1, \ldots, T\}$ be the probability that AdaBoost assigns to the i'th instance of the training data after the t'th round. Compute $\sum_{i:h_t(x_i)=y_i} p_i^{(t+1)}$, i.e., the updated total weight of all instances that are correctly classified by h_t — the hypothesis chosen by AdaBoost in the t'th round.

II. Neural networks

1. Let $\mathcal{F} \subseteq \{-1, 1\}^{\mathbb{R}^d}$ be the class of functions that correspond to single Perceptron-like artificial neurons with activation function $\sigma(z) = \mathrm{sgn}(z)$ and d inputs. Let AdaBoost run on this set as set of base hypotheses for T rounds. Can you represent the resulting function in terms of a neural network? What can you say about the number of hidden neurons, layers and about the weights?

2. Compare the neural network in your brain to present day computers—quantitatively as well as qualitatively. What are similarities, what are differences?

3. Let $A := \{x_1, \ldots, x_N\} \subseteq \mathbb{R}^d$ be N points in general position (i.e., no hyperplane contains more than d of those points). Show that for every $f : A \to \{0, 1\}$ there is a neural network with a single hidden layer of at most $\lceil N/d \rceil$ neurons and with activation function $\sigma(z) = \mathbb{1}_{z \geq 0}$ that implements a function $F : \mathbb{R}^d \to \{0, 1\}$ so that $F|_A = f$.