Prof. Dr. M. Wolf
J. Cuesta

TECHNISCHE UNIVERSITÄT MÜNCHEN

Zentrum Mathematik

Mathematical Foundations of Machine Learning

http://www-m5.ma.tum.de/Allgemeines/MA4801_2018S

Summer 2018
Blatt 8
(06.06.2018)

The following homework will be discussed on June 12th and June 15th:

H8.1. (VC-dimension of linearly combined classifiers)

a) Let d be the VC-dimension of $\mathcal{F} \subset \{-1,1\}^{\mathcal{X}}$ and for $T \in \mathbb{N}$ define $\mathcal{F}_T := \{f = \operatorname{sgn} \sum_{t=1}^{T} w_t h_t \mid w_t \in \mathbb{R},\ h_t \in \mathcal{F}\}$. Show that the growth function Γ of $\mathcal{F}_T$ satisfies

$$\Gamma(n) \leq \left(\frac{en}{T}\right)^T \left(\frac{en}{d}\right)^{Td}$$

b) Provide an upper bound on the VC-dimemsion of $\mathcal{F}_T$ for the case where F is the set of indicator functions for closed balls in $\mathbb{R}^n$. (See exercise H2.5)

H8.2. (**VC-dimension for step-activation functions**)

a) Let $n_0, \omega \in \mathbb{N}$ and fix an architecture of a multilayered feedforward neural network with n_0 inputs, a single output and ω parameters (i.e. weights and threshold values). Denote by $\mathcal{F}$ the set of all functions $f : \mathbb{R}^{n_0} \to \{-1,1\}$ that can be implemented by any feedforward network with this architecture when using $\sigma(z) = \operatorname{sgn}(z)$ as activation function. Show that

$$\operatorname{VCdim}(\mathcal{F}) < 2\omega \log_2(e\omega).$$

b) The Large Scale Visual Recognition Challenge (ILSVRC) is a contest that aims to estimate the content of photographs using as training data a subset of the hand-labeled ImageNet dataset (10,000,000 labeled images depicting 10,000+ object categories). In 2012 Alex Krizhevsky, I. Sutskever and G. E. Hinton made a submission using a neural network which won the competition and made an influential contribution to the field of computer vision models. Look for this paper and compute an upper bound for the VCdim of this neural network assuming that the output would be binary.

H8.3. (**Approximation of $C([0,1]^n)$ by continuous sigmoids**) Let $C([0,1]^n)$ be the space of real-valued continuous functions with support on the $n-$dimensional unit cube. A *sigmoid* is a bounded function $\sigma : \mathbb{R} \to \mathbb{R}$ that satisfies $\lim_{z\to\infty} \sigma(z) = 1$ and $\lim_{z\to-\infty} \sigma(z) = 0$. We say that σ is *discriminatory* if for a signed regular Borel measure μ on $[0,1]^n$

$$\int_{[0,1]^n} \sigma(\langle y, x\rangle + \theta) d\mu(x) = 0$$

for all $y \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ implies $\mu = 0$.

a) Show that the set

$$S = \left\{ g : [0,1]^n \to \mathbb{R} \;\middle|\; \exists k \in \mathbb{N}, y \in \mathbb{R}^n, \theta \in \mathbb{R}^k, \alpha \in \mathbb{R}^k : \ g(x) = \sum_{j=1}^{k} \alpha_j \sigma(\langle y, x\rangle + \theta_j) \right\}$$

is dense in $C([0,1]^n)$ with respect to the supremum (or uniform) norm $\|f\| = \sup_{x\in[0,1]^n} |f(x)|$ if σ is discriminatory. In other words, given any $f \in C([0,1]^n)$ and $\epsilon > 0$, there exist a $g \in S$ with σ discriminatory for which $|f(x) - g(x)| < \epsilon$ for all $x \in [0,1]^n$.

Hint: Look for an appropiate version of the Hahn-Banach theorem and Riesz Representation theorem. See for instance Chapter 3 in the book "W. Rudin, Functional Analysis, McGraw-Hill (1973)".

b) (Optional) Show that any sigmoid, as defined here, is discriminatory.

H8.4. (**Approximation by exponentials**) Let $K \subset \mathbb{R}^n$ be compact. Prove that

$$S := \operatorname{span}\{f : K \to \mathbb{R} \mid f(x) = \exp \sum_{i=1}^{n} w_i x_i,\ w_i \in \mathbb{R}\}$$

is dense in $C(K)$ with respect to the uniform norm.
Hint: Look at for the statement of the Stone-Weierstrass theorem.