TECHNISCHE UNIVERSITÄT MÜNCHEN
Zentrum Mathematik

PROF. DR. M. WOLF **Mathematical Foundations of Machine Learning** Summer 2016

Blatt 1

http://www-m5.ma.tum.de/Allgemeines/MA4801_2016S (13.04.2016)

The following homework will be discussed on Wed. April 20th:

1. **Empirical risk minimization**

   (a) Let $\mathcal{X} \times \mathcal{Y} = \mathbb{R}^d \times \mathbb{R}$ and $\mathcal{F} := \{h : \mathcal{X} \to \mathcal{Y} | \exists v \in \mathbb{R}^d : h(x) = \langle v, x \rangle\}$ be the class of linear functions. For $S = \big((x_i, y_i)\big)_{i=1}^n \in (\mathcal{X} \times \mathcal{Y})^n$ and the quadratic loss, derive the hypothesis $\hat{h} \in \mathcal{F}$ that minimizes the empirical risk $\hat{R}$.

   (b) Let $\mathcal{X} \times \mathcal{Y} = \mathbb{R} \times \mathbb{R}$ and $\mathcal{F} := \{h : \mathcal{X} \to \mathcal{Y} | \exists a \in \mathbb{R}^{m+1} : h(x) = \sum_{k=0}^m a_k x^k\}$ the set of all polynomials of degree m. Use (a) to derive $\hat{h} \in \mathcal{F}$ that minimizes the empirical risk for a given set $S \in (\mathcal{X} \times \mathcal{Y})^n$, again w.r.t. the quadratic loss.

2. **Error decomposition**
   In the setup of the first Thm. of the lecture, consider a fixed learner that outputs a hypothesis h_S upon input of $S \in (\mathcal{X} \times \mathcal{Y})^n$. Regard S as a random variable, distributed according to P^n and define $\bar{h}(x) := \mathbb{E}_S[h_S(x)]$ the expected prediction for a fixed x. Prove that if the expected risk $\mathbb{E}_S[R(h_S)]$ is finite, then it is equal to

   $$\mathbb{E}\left[|Y - r(X)|^2\right] + \mathbb{E}\left[|\bar{h}(X) - r(X)|^2\right] + \mathbb{E}\left[\mathbb{E}_S\left[|h_S(X) - \bar{h}(X)|^2\right]\right]. \quad (1)$$

3. **Hoeffding's inequality**
   Let S, S' be i.i.d. random variables, each with values in $(\mathcal{X} \times \mathcal{Y})^n$ and distributed according to some product measure P^n. Consider a loss function whose range is contained in a finite interval of length $c > 0$ and denote by $\hat{R}(h)$ and $\hat{R}'(h)$ the empirical risks of a fixed hypothesis $h \in \mathcal{Y}^{\mathcal{X}}$ w.r.t. S and S' respectively. Derive an upper bound for the probability $\mathbb{P}_{SS'}\left[|\hat{R}(h) - \hat{R}'(h)| > \epsilon\right]$ that depends on n, ϵ and c.
   (You may use Hoeffding's inequality for that. If necessary, look it up.)