Winter term 2024/25
U. von Luxburg
E. Günther/ K. Frohnapfel

Assignment 11

Mathematics for Machine Learning

Submission due Friday **17.01.2025, 23:59** via Ilias

Justify all your claims.

Exercise 1 (**Convergence of random variables**, 3+2+2 points)**.**

a) Consider the probability space $([0,1], \mathcal{B}, \lambda)$, where $\mathcal{B}$ denotes the Borel-σ algebra on $[0,1]$ and λ the Lebesgue measure. For every $n \in \mathbb{N}$ there exist unique $h, k \in \mathbb{N}_0$ with $0 \leq k < 2^h$ such that $n = 2^h + k$. We define a sequence of random variables X_n using these h, k as

$$X_n(\omega) = \mathbb{1}_{\left[\frac{k}{2^h}, \frac{k+1}{2^h}\right]}(\omega) = \begin{cases} 1, & \omega \in \left[\frac{k}{2^h}, \frac{k+1}{2^h}\right] \\ 0, & \text{otherwise} \end{cases} \qquad \forall \omega \in [0,1]\,.$$

Prove that $X_n \to 0$ as $n \to \infty$ in probability and in L^1, but not almost surely.

b) Consider a sequence of random variables $X, X_1, X_2, \ldots$ that satisfies $X_n \to X$ in probability as $n \to \infty$. Prove that there exists a subsequence $(X_{n_k})_{k \in \mathbb{N}}$ for which $X_{n_k} \to X$ almost surely as $k \to \infty$.

c) Find such a subsequence for the sequence given in a).

Exercise 2 (**Limit theorems**, 2+3 points)**.**

a) A laundry bag contains one black and one white sock. Now Tom keeps throwing socks into the laundry bag. Every sock he throws is either black with probability $p \in [0,1)$ or white with probability $1-p$, independently of the previous socks. Let X_n be the fraction of black socks to total amount of socks and Y_n the fraction of black to white socks after $n \in \mathbb{N}$ throws. Prove that

   (a) $X_n \to p$ almost surely as $n \to \infty$,

   (b) $Y_n \to \frac{p}{1-p}$ almost surely as $n \to \infty$.

b) Consider an i.i.d. sequence of real-valued random variables $(X_n)_{n \in \mathbb{N}}$ with $\mu = \mathbb{E}[X_1] \in \mathbb{R}$ and $\sigma^2 = \mathrm{Var}[X_1] < \infty$. Define $S_n \coloneqq \sum_{k=1}^n X_k$ and let $a, b \in \mathbb{R}$ with $a < b$. Use the central limit theorem to prove

$$P(a \leq S_n \leq b) = \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right) + o(1)\,,$$

where Φ denotes the cumulative distribution function (cdf) of the standard normal distribution and $o(1)$ satisfies $o(1) \to 0$ as $n \to \infty$.
Hint: You may use the characterization

$$X_n \xrightarrow[n\to\infty]{} X \text{ in distribution} \quad \Leftrightarrow \quad F_n \xrightarrow[n\to\infty]{} F \text{ uniformly on } D_F\,,$$

where F_n and F denote the cdf's of X_n and X, and $D_F = \{x \in \mathbb{R} \mid F \text{ is continuous at } x\}$.

Exercise 3. (**Estimation error in learning theory**, 3+2+2+1 points)
Let $(X,Y),(X_1,Y_1),...,(X_n,Y_n)$ be i.i.d. random variables $\Omega \to \mathbb{R}^d \times \mathbb{R}$ and consider a finite set $\mathcal{H}$ of measurable prediction functions $h : \mathbb{R}^d \to \mathbb{R}$.
For a measurable loss function $l : \mathbb{R} \times \mathbb{R} \to [0,1]$ we define the empirical risk of a predictor h as

$$R_n(h) := \frac{1}{n} \sum_{i=1}^{n} l(h(X_i), Y_i)$$

and its true risk by

$$R(h) := \mathrm{E}(l(h(X), Y)).$$

a) Let $\varepsilon > 0$. Prove that for the event $A_n := \{\sup_{h \in \mathcal{H}} |R_n(h) - R(h)| \geq \varepsilon\}$ it holds

$$P(A_n) \leq M \cdot \mathrm{e}^{-2n\varepsilon^2}$$

for some constant $M > 0$.

b) Let $h \in \mathcal{H}$ be fixed. Prove that $R_n(h) \to R(h)$ almost surely as $n \to \infty$.

c) Define

$$h_n := \arg\min_{h \in \mathcal{H}} R_n(h) \quad \text{and} \quad h^* := \arg\min_{h \in \mathcal{H}} R(h)\,.$$

Prove that $R(h_n) \to R(h^*)$ almost surely as $n \to \infty$.
Hint: Show and use that

$$R(h_n) - R(h^*) \leq 2 \sup_{h \in \mathcal{H}} |R_n(h) - R(h)|\,.$$

d) The result of c) does not hold any more in general if $\mathcal{H}$ is not finite. Explain why.