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

Assignment 09

Mathematics for Machine Learning

Submission due Friday **20.12.2024, 23:59** via Ilias

Justify all your claims.

Exercise 1 (**σ-Algebra**, 3 points)**.**
Let X be a countable set and $\mathcal{A}$ a σ-Algebra on X, which separates points, that is, for any $x_1, x_2 \in X$ with $x_1 \neq x_2$ there exists a set $A \in \mathcal{A}$ such that $x_1 \in A$ and $x_2 \notin A$. Show that $\mathcal{A} = \mathcal{P}(X)$, where $\mathcal{P}(X)$ is the power set of X.
Hint: Show and use that a σ-Algebra is closed under countable intersection.

Exercise 2 (**Covariance and correlation**, 2+1+3 points)**.**
Consider the probability space $(\Omega, \mathcal{A}, P)$, the measurable spaces $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ and two random variables $X, Y \in L^2(\Omega, \mathcal{A}, P)$ with $X : \Omega \to \Omega_1$ and $Y : \Omega \to \Omega_2$.

a) Consider two measurable functions $f : \Omega_1 \to \mathcal{W}_1$ and $g : \Omega_2 \to \mathcal{W}_2$ for the measurable spaces $(\mathcal{W}_1, \mathcal{F}_1)$ and $(\mathcal{W}_2, \mathcal{F}_2)$.
Show that the independence of X and Y implies the independence of $f(X)$ and $g(Y)$.

b) Prove that the correlation coefficient satisfies $\varrho_{X,Y} \in [-1, 1]$.

c) Consider the following concrete setting of two independent dice throws:

$$\Omega = \Omega_1 = \Omega_2 = \{1, \dots, 6\}$$
$$\mathcal{A} = \mathcal{A}_1 = \mathcal{A}_2 = \mathcal{P}(\Omega)$$
$$P(X = k) = P(Y = k) = \frac{1}{6} \text{ for all } i \in \{1, \dots, 6\}$$

Assume that X and Y are independent and compute the correlation coefficient ϱ_{W_1,W_2} for $W_1 := X + Y$ and $W_2 := X \cdot Y$.

Exercise 3 (**Bernoulli trials**, 2+2+1 points)**.**
A Bernoulli trial is a random experiment with only two outcomes, "success" and "failure". Assume that the probability of success is $p \in (0, 1)$. This Bernoulli trial is now repeated arbitrarily often, where the outcome of each trial is independent of the other trials.
For $n \in \mathbb{N}$, let K_n denote the number of successes within the first n trials, T_1 the number of failures until the first success, and T_2 the total number of failures until the second success.

a) Prove that $K_n \overset{d}{=} B(n, p)$ for any $n \in \mathbb{N}$. This means that K_n is distributed as the binomial distribution $B(n, p)$, which is defined by $B(n, p)(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k \in \{0, \dots, n\}$.

b) Compute the distribution of T_1, T_2, and $X \coloneqq T_2 - T_1$.

c) Are X and T_1 independent? (Prove your answer.)

Exercise 4 (**Independence**, 4+2 points)**.**
Consider the probability space $(\Omega, \mathcal{B}, P)$ for the Borel-σ-Algebra $\mathcal{B}$ and a real valued random variable $X : \Omega \to \mathbb{R}$.

a) Show that X is independent of itself if and only if it is almost surely constant.
Hint: Use that $P(X = c) = P(X \leq c) - P(X < c)$ *for a constant* $c \in \mathbb{R}$.

b) Deduce from a) that X is independent of $f(X)$ for a measurable function f if and only if $f(X)$ is almost surely constant.
Hint: You can use all other exercises on this sheet.