Optimization in Machine Learning **Universität Mannheim**
HWS 2026 Prof. Leif Döring, Benedikt Wille

Solution Sheet 4

For the exercise class on the 23.11.2026 at 12:00.

Exercises to be solved in the exercise class

Exercise 1 (Conditional Expectation)**.** **(0 Points)**

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{F}$ a subalgebra of $\mathcal{A}$ and X, Y random vectors. Prove for $\mathcal{F}$-measurable $X \in \mathbb{R}^d$ that we have

$$\mathbb{E}[\langle X, Y\rangle \,|\, \mathcal{F}] = \langle X, \mathbb{E}[Y \,|\, \mathcal{F}]\rangle$$

Solution. We simply use linearity of the conditional expectation

$$\mathbb{E}[\langle X, Y\rangle \,|\, \mathcal{F}] = \mathbb{E}\left[\sum_{i=1}^d X_i Y_i \,|\, \mathcal{F}\right] = \sum_{i=1}^d \mathbb{E}\big[X_i Y_i \,|\, \mathcal{F}\big] = \sum_{i=1}^d X_i \mathbb{E}\big[Y_i \,|\, \mathcal{F}\big] = \langle X, \mathbb{E}[Y \,|\, \mathcal{F}]\rangle. \quad \square$$

Exercise 2 (Conditional expectation)**.** **(0 Points)**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an underlying probability space.

(i) Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra and let $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Prove that

(a) for any $\lambda \in \mathbb{R}$ there holds $\mathbb{E}[\lambda X + Y \,|\, \mathcal{G}] = \lambda \mathbb{E}[X \,|\, \mathcal{G}] + \mathbb{E}[Y \,|\, \mathcal{G}]$.

Solution. Let $A \in \mathcal{G}$. Then

$$\begin{aligned}\mathbb{E}\big[\mathbb{1}_A\big(\lambda \mathbb{E}[X \,|\, \mathcal{G}] + \mathbb{E}[Y \,|\, \mathcal{G}]\big)\big] &= \lambda \mathbb{E}\big[\mathbb{1}_A \mathbb{E}[X \,|\, \mathcal{G}]\big] + \mathbb{E}\big[\mathbb{1}_A \mathbb{E}[Y \,|\, \mathcal{G}]\big] \\ &= \lambda \mathbb{E}\big[\mathbb{1}_A X\big] + \mathbb{E}\big[\mathbb{1}_A Y\big] \\ &= \mathbb{E}\big[\mathbb{1}_A(\lambda X + Y)\big]\end{aligned}$$

Thus, by uniqueness of the conditional expectation, $\mathbb{E}[\lambda X + Y \,|\, \mathcal{G}] = \lambda \mathbb{E}[X \,|\, \mathcal{G}] + \mathbb{E}[Y \,|\, \mathcal{G}]$ almost surely. $\square$

(b) if $X \geq Y$ $\mathbb{P}$-almost surely, then $\mathbb{E}[X \,|\, \mathcal{G}] \geq \mathbb{E}[Y \,|\, \mathcal{G}]$ $\mathbb{P}$-a.s. .

Solution. Let $A := \{\mathbb{E}[X \,|\, \mathcal{G}] < \mathbb{E}[Y \,|\, \mathcal{G}]\}$. Then A is $\mathcal{G}$ measurable and it follows from the definition of the conditional expectation that

$$0 \overset{\text{def. } A}{\geq} \mathbb{E}[\mathbb{1}_A\big(\mathbb{E}[X \,|\, \mathcal{G}] - \mathbb{E}[Y \,|\, \mathcal{G}]\big)] = \mathbb{E}[\mathbb{1}_A(X - Y)] \overset{X \geq Y}{\geq} 0.$$

Thus, $\mathbb{1}_A(X - Y) = 0$ almost surely and by definition of A this implies $\mathbb{1}_A = 0$ almost surely, i.e. $\mathbb{E}[X \,|\, \mathcal{G}] \geq \mathbb{E}[Y \,|\, \mathcal{G}]$ almost surely. $\square$

(c) $|\mathbb{E}[X \,|\, \mathcal{G}]| \leq \mathbb{E}[|X| \,|\, \mathcal{G}]$.

Solution. One possibility is to apply Jensen's inequality for the conditional expectation to the convex function $x \mapsto |x|$. Another way is to split the random variable X into its positive and negative part $X = X^+ - X^-$ and to use the linearity of the conditional expectation and the previous result, i.e. $X^+ \geq 0$ almost surely implies $\mathbb{E}[X^+ \,|\, \mathcal{G}] \geq 0$ almost surely. □

(ii) Let Y_1,Y_2 be iid. random variables with

$$\mathbb{P}(Y_1 = 2) = \mathbb{P}(Y_1 = 0.5) = \frac{1}{2}\,,$$

and set

$$S_0 = 2, \quad S_k = S_0 \cdot \prod_{i=1}^{k} Y_i\,,\ k = 1, 2\,.$$

Compute $\mathbb{E}[S_2 \,|\, \mathcal{F}_1]$, where $\mathcal{F}_1 = \sigma(S_1)$.

Solution. We have

$$\mathbb{E}[S_2 \,|\, \mathcal{F}_1] = \mathbb{E}[S_1 Y_2 \,|\, \mathcal{F}_1] = S_1 \mathbb{E}[Y_2 \,|\, \mathcal{F}_1] = S_1 \mathbb{E}[Y_2] = S_1\big(\tfrac{2}{2} + \tfrac{1/2}{2}\big) = \tfrac{5}{2} S_1. \qquad \square$$

(iii) Let X, Y be independent Bernoulli distributed random variables with parameter $p \in [0, 1]$ and define $Z := \mathbb{1}_{\{X+Y=0\}}$.

(a) Compute $\mathbb{E}[X \,|\, \sigma(Z)]$ and $\mathbb{E}[Y \,|\, \sigma(Z)]$.

Solution. We have

$$\mathbb{P}(Z = 1) = \mathbb{P}(X + Y = 0) = \mathbb{P}(X = 0, Y = 0) = (1 - p)^2. \tag{1}$$

Thus,

$$\begin{aligned}
\mathbb{E}[X \,|\, \sigma(Z)] &= \mathbb{E}[X \,|\, Z = 0]\mathbb{1}_{Z=0} + \underbrace{\mathbb{E}[X \,|\, Z = 1]}_{=0} \mathbb{1}_{Z=1} \\
&= \frac{\mathbb{E}[X \mathbb{1}_{Z=0}]}{\mathbb{P}(Z = 0)} \mathbb{1}_{Z=0} \\
&= \frac{0\mathbb{P}(X = 0, Y = 1) + 1[\mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 1, Y = 1)]}{1 - (1 - p)^2} \mathbb{1}_{Z=0} \\
&= \frac{p(1 - p) + p^2}{1 - (1 - 2p + p^2)} \mathbb{1}_{Z=0} \\
&= \frac{p}{2p - p^2} \mathbb{1}_{Z=0} \\
&= \frac{1}{2 - p} \mathbb{1}_{Z=0}.
\end{aligned}$$

By a symmetric argument we have $\mathbb{E}[Y \,|\, \sigma(Z)] = \frac{1}{2-p} \mathbb{1}_{Z=0}$. □

(b) When are these random variables independent?

Solution. Since we have $\mathbb{E}[X \,|\, \sigma(Z)] = \mathbb{E}[Y \,|\, \sigma(Z)]$ almost surely by the previous result, $\mathbb{E}[X \,|\, \sigma(Z)]$ is independent from $\mathbb{E}[Y \,|\, \sigma(Z)]$ if and only if it is a constant almost surely. For this to be the case, we need $\mathbb{1}_{Z=0}$ to be constant almost surely, i.e. $\mathbb{P}(Z = 0) = 0$ or $\mathbb{P}(Z = 0) = 1$. This is only the case for $p \in \{0, 1\}$ by (1).
In summary, $\mathbb{E}[X \,|\, \sigma(Z)]$ is independent from $\mathbb{E}[Y \,|\, \sigma(Z)]$ if and only if $p \in \{0, 1\}$. □