# Mathematik des maschinellen Lernens (WiSe 2021/22)

## 2. Übung

03. November 2021

**1. Aufgabe**$^{(*)}$

Es sei $\mathcal{X}$ beliebig und $\mathcal{Y} = \{0, 1\}$ sowie $\ell(y, y') = 1 - \delta_{yy'}$ der 0-1-Verlust. Dieser unterliegt einer Bernoulli-Verteilung $B(p)$, $p \in [0, 1]$, denn für $L = \ell(h(X), Y)$ mit $(X, Y) \sim \mathcal{D}$ gilt:

$$L = \begin{cases} 1 & \text{mit Wahrscheinlichkeit p,} \\ 0 & \text{mit Wahrscheinlichkeit 1-p.} \end{cases}$$

a) Von welchen Faktoren ist die Wahrscheinlichkeit $p$ alles abhängig?

b) Interpretieren Sie den Erwartungswert von $L$.

c) Berechnen Sie die Varianz von $L$.

**2. Aufgabe**$^{(**)}$

Wir betrachten folgende Aufgabe:

- Auf Basis des monatlichen Bruttoeinkommens $x$ wollen wir vorhersagen, ob ein Kredit von 200000 Euro innerhalb von 15 Jahren vollständig zurückgezahlt wird (Markierung: $y = 0$) oder ob er ausfällt (Markierung: $y = 1$).
- Dazu machen wir folgende Annahme: Unterschreitet das Bruttoeinkommen eine gewisse Schranke, z. B. 4000 Euro, so wird der Kredit ausfallen.
- Wir gehen dabei von 10 möglichen Schranken für das Bruttomonatseinkommen aus (in Euro): 4400, 4600, 4800, ..., 6200.
- Wir wollen nun die Schranke lernen, die möglichst wenig falsche Vorhersagen des Kreditausfalls in der Praxis erzielt.

a) Formulieren Sie die Hypothesenklasse $\mathcal{H}$ in dieser Aufgabe. Welche Verlustfunktion $\ell$ würden Sie hier wählen?

b) Angenommen, das Bruttoeinkommen $X$ in der repräsentativen Bevölkerungsschicht von KreditnehmerInnen ist gleichverteilt zwischen 3000 Euro und 7000 Euro und es gilt, dass (fast sicher) der Kredit vollständig zurückgezahlt wird, falls das Bruttoeinkommen mindestens 5400 Euro beträgt.

Berechnen Sie für alle zehn Hypothesen $h$ in der Hypothesenklasse $\mathcal{H}$ das resultierende erwartete Risiko $\mathcal{R}_{\mathcal{D}}(h)$.

c) Erfüllt die Hypothesenklasse $\mathcal{H}$ in diesem Beispiel die Realisierbarkeitsannahme?

d) Welche Hypothesen aus der Hypothesenklasse erzielen eine Falschklassifikationsrate bzw. -wahrscheinlichkeit von maximal 10%?

e) Wir wollen nun erreichen, dass wir per (verzerrter) ERM-Lernregel mit einer Wahrscheinlichkeit von mindestens 95% eine Hypothese (bzw. Schranke) lernen, die maximal 10% Falschvorhersagen liefert.

Formulieren Sie die entsprechende PAC-Bedingung und schätzen Sie mittels Proposition 2.4 der Vorlesung die Mindestanzahl $m$ von Trainingsdaten (Kreditdaten) ab, um dieses Lernziel zu erreichen.

f) Überprüfen Sie Ihre Abschätzung für $m$ mit Hilfe des zur Verfügung gestellten Matlab-Codes `kredit_sim.m`, welcher 1000 zufällige Trainingsdatensätze jeweils im Umfang von $m$ Daten generiert und für diese jeweils die ERM-Regel anwendet. Setzen Sie dazu in Zeile

23 Ihren Wert für $m$ ein und überprüfen Sie mittels Zeile 76, wie oft die entsprechenden Hypothesen aus d) erlernt wurden.
Wählen Sie für $m$ im Matlabcode auch den halben Wert Ihrer Schranke. Was beobachten Sie?

g) Wie ändert sich Ihre Abschätzung für $m$, falls die Einkommen anders verteilt sind, z. B. gleichverteilt zwischen 1000 und 8000 Euro? Für welche Einkommensverteilungen gilt die Abschätzung für $m$ unter der Annahme, dass der Kredit bei mehr als 5400 Euro stets zurückgezahlt wird?

**3. Aufgabe**$^{(*)}$ (Lernbarkeit)
Weisen Sie formal nach, dass eine bzgl. eines Verlusts $\ell$ (agnostisch) PAC-lernbare Klasse $\mathcal{H}$ insbesondere bzgl. $\ell$ PAC-lernbar unter der Realisierbarkeitsannahme ist.

**4. Aufgabe**$^{(*)}$ (Informationskomplexität von $\mathcal{H}$)
Ist eine Klasse $\mathcal{H}$ bzgl. eines Verlustes $\ell$ durch einen Lernalgorithmus $A$ PAC-lernbar – egal ob agnostisch oder ob unter der Realisierbarkeitsannahme – dann gibt es viele Funktionen $m_{\mathcal{H}}$ für hinreichende Stichprobenumfänge bzgl. der jeweiligen PAC-Bedingung. Die Funktion $m_{\mathcal{H}}$, die der PAC-Lernbarkeit genügt und für jedes $\epsilon, \delta \in (0,1)$ jeweils die kleinstmöglichen Werte in $\mathbb{N}$ annimt, nennt man *Informationskomplexität von $\mathcal{H}$*.

Zeigen Sie, dass die Informationskomplexität

$$m_{\mathcal{H}}\colon (0,1)^2 \to \mathbb{N},\ m_{\mathcal{H}}(\epsilon,\delta) = \left\lceil \frac{\ln(|\mathcal{H}|/\delta)}{\epsilon} \right\rceil$$

aus Proposition 2.4 in beiden Argumenten $\epsilon, \delta \in (0,1)$ monoton nicht wachsend ist.