

**Algorithmische Grundlagen des Maschinellen Lernens**  
Sommersemester 2020  
Übungsblatt 2

**Aufgabe 1:** (5 Punkte)

(Vereinigung von  $k$  Intervallen) Betrachten Sie als Hypothesenklasse  $\mathcal{H}$  die Menge aller Funktionen der Form

$$h_{a,b}: \mathbb{R} \rightarrow \{-1, +1\}, \text{ mit } a, b \in \mathbb{R}^k \text{ und}$$
$$h_{a,b}(x) = \begin{cases} +1 & \text{falls } x \in \bigcup_{i=1}^k [a_i, b_i] \\ -1 & \text{sonst} \end{cases}$$

Sei  $S = \{x_1, \dots, x_m\}$ . Wieviele verschiedene Wege gibt es, den Elementen in  $S$  Labels zu geben mithilfe einer Funktion in  $\mathcal{H}$ ? In anderen Worten, wie groß ist  $|\mathcal{H}|_S$ ?

**Aufgabe 2:** (5 Punkte)

Sei die Grundmenge  $\mathbb{R}^2$ . Zeigen Sie

1. Das Mengensystem aller achsenparallelen Quadrate hat VC-Dimension höchstens 3.
2. Das Mengensystem aller Quadrate hat VC-Dimension mindestens 4.

**Aufgabe 3:** (3 Punkte)

Sei  $\mathcal{R}$  ein Mengensystem mit  $m$ -elementiger Grundmenge und  $\dim(\mathcal{R}) = 0$ .

1. Zeigen Sie  $|\mathcal{R}| \leq 1$ .
2. Gibt es eine obere Schranke für  $m$  in diesem Fall?

**Aufgabe 4:** (3 Punkte)

Sei  $\mathcal{R}$  ein Mengensystem mit Grundmenge  $\mathcal{X}$  und sei  $A \subseteq \mathcal{X}$  eine beliebige Teilmenge. Zeigen Sie, dass  $\dim(\mathcal{R}|_A) \leq \dim(\mathcal{R})$ .

**Aufgabe 5:** (4 Punkte)

Es sei  $\mathcal{D}$  eine diskrete Wahrscheinlichkeitsverteilung auf einer endlichen Menge. Es gibt also  $a_1, \dots, a_n$ , sodass  $\Pr_{x \sim \mathcal{D}} [x = a_i] = p_i$  und  $\sum_{i=1}^n p_i = 1$ .

Zeigen Sie, dass die folgenden zwei Wege, eine Menge  $S$  und  $S'$  zu erzeugen, äquivalent sind.

1.  $S$  ist eine Menge von  $m$  zufällig aus  $\mathcal{D}$  gezogenen Samples.  $S'$  ist eine Menge von  $m$  zufällig aus  $\mathcal{D}$  gezogenen Samples.
2. Es seien  $X_1, \dots, X_{2m}$  unabhängige, identisch verteilte Züge aus  $\mathcal{D}$ . Nun ziehe  $m$  mal zufällig gleichverteilt ohne Zurücklegen aus diesen Zügen und nenne das Ergebnis  $S$ . Die Menge  $S'$  umfasst alle verbleibenden Elemente.

Tipp: Vergleichen Sie die Wahrscheinlichkeiten, dass ein Element  $a_i$  (nicht) in der Menge enthalten ist.