Computational Learning Theory

Sommersemester 2017

Mario Holldack, M. Sc.
Prof. Dr. Georg Schnitger
Hannes Seiwert, M. Sc.

GOETHE UNIVERSITÄT FRANKFURT AM MAIN
Institut für Informatik
AG Theoretische Informatik

Übungsblatt 3

Ausgabe: 08.05.2017
Abgabe: 15.05.2017

Aufgabe 3.1 *VC-Dimension der Sinus-Klasse* (8 Punkte)

Für jedes $a \in \mathbb{R}$ sei die charakteristische Funktion $\mathbb{1}_a : \mathbb{R} \to \{0,1\}$ durch

$$\mathbb{1}_a(x) := \begin{cases} 1, & \text{falls } \sin(ax) > 0 \\ 0, & \text{falls } \sin(ax) \leq 0 \end{cases}$$

gegeben. Die Konzeptklasse $\mathfrak{S}$ enthalte alle Konzepte $\mathsf{Sinus}(a) := \mathbb{1}_a^{-1}(1) = \{x \in \mathbb{R} \; : \; \mathbb{1}_a(x) = 1\}$. Zeigen Sie: $\mathrm{VC}(\mathfrak{S}) = \infty$.

Hinweis: Sei $z_2(x) = 0, b_1 b_2 b_3 \dots$ die Binärdarstellung einer reellen Zahl $x \in (0,1)$, wobei $z_2(x)$ unendlich viele Einsen enthalte. Verwenden Sie (ohne Beweis): $\mathbb{1}_{2^m \pi}(x) = 1 - b_m$ für jedes $m \in \mathbb{N}_{>0}$.

Fazit: Obwohl jedes Konzept $\mathsf{Sinus}(a)$ nur durch einen Parameter a charakterisiert wird, ist die VC-Dimension unendlich. Somit gibt es für $\mathfrak{S}$ keinen PAC-Algorithmus.

Aufgabe 3.2 *Eigenschaften der VC-Dimension* (12 Punkte)

Sei X ein Beispielraum.

a) Seien $\mathcal{C}_1, \mathcal{C}_2$ und $\mathcal{C}$ Konzeptklassen über X.

   i) Sei $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Zeigen Sie: $\mathrm{VC}(\mathcal{C}_1) \leq \mathrm{VC}(\mathcal{C}_2)$.

   ii) Sei $\overline{\mathcal{C}} := \{X \setminus c : c \in \mathcal{C}\}$ die Komplement-Konzeptklasse von $\mathcal{C}$. Zeigen Sie: $\mathrm{VC}\left(\overline{\mathcal{C}}\right) = \mathrm{VC}(\mathcal{C})$.

   iii) Sei $\mathcal{C}_1 \vee \mathcal{C}_2 := \{c_1 \cup c_2 : c_1 \in \mathcal{C}_1, c_2 \in \mathcal{C}_2\}$ die elementweise Vereinigung von $\mathcal{C}_1$ und $\mathcal{C}_2$. Widerlegen Sie die folgende Ungleichung: $\mathrm{VC}(\mathcal{C}_1 \vee \mathcal{C}_2) \leq \mathrm{VC}(\mathcal{C}_1) + \mathrm{VC}(\mathcal{C}_2)$.

   Hinweis: Sie können die Konzeptklassen $\textsc{Kugel}_2$ und $\textsc{Gerade}_2$ aller Kreisscheiben und Geraden in der reellen Ebene als Gegenbeispiel verwenden.

b) Sei $r \in \mathbb{N}_{>0}$. Seien $\mathcal{C}_1, \dots, \mathcal{C}_r$ Konzeptklassen über X, sei $\mathcal{C} := \mathcal{C}_1 \cup \dots \cup \mathcal{C}_r$ ihre Vereinigung und sei $d := \max\left\{\mathrm{VC}(\mathcal{C}_i) : i \in \{1, \dots, r\}\right\} \geq 3$. Zeigen Sie:

$$\mathrm{VC}(\mathcal{C}) \leq 4d \ln(2d) + 2\ln(r)$$

Hinweis: Sei S eine Beispielmenge. Schätzen Sie $\Pi_{\mathcal{C}}(S)$ durch $|\Pi_{\mathcal{C}_1}(S)| + \dots + |\Pi_{\mathcal{C}_r}(S)|$ ab und verwenden Sie anschließend Sauers Lemma. Sie erhalten auch die volle Punktzahl, wenn Sie die folgende Ungleichung zeigen:

$$2^{\mathrm{VC}(\mathcal{C})} \leq r \cdot \left(\frac{e \cdot \mathrm{VC}(\mathcal{C})}{d}\right)^d.$$

Bitte wenden!

Aufgabe 3.3 *Intervalle lernen* (12 Punkte + 6 Bonuspunkte)

Sei $\textsc{Intervall} = \{[a,b] : a,b \in \mathbb{R},\ a \leq b\}$ die Konzeptklasse der abgeschlossenen reellen Intervalle. Für jedes $k \in \mathbb{N}_{>0}$ definieren wir

$$\textsc{Intervall}^k := \bigvee_{i=1}^{k} \textsc{Intervall} = \left\{ \bigcup_{i=1}^{k} [a_i, b_i] \ : \ [a_1, b_1], \ldots, [a_k, b_k] \in \textsc{Intervall} \right\}$$

als die Konzeptklasse aller Vereinigungen von höchstens k Intervallen. Schließlich sei

$$\textsc{Intervall}^+ := \bigcup_{i=1}^{\infty} \textsc{Intervall}^i$$

die Konzeptklasse aller Vereinigungen endlich vieler Intervalle.

a) Zeigen Sie: $\mathrm{VC}\left(\textsc{Intervall}^k\right) = 2k$ für alle $k \in \mathbb{N}_{>0}$.

b) Zeigen Sie: Es gibt keinen PAC-Algorithmus für $\textsc{Intervall}^+$.

c) Entwerfen Sie einen Lernalgorithmus, der für jede Verteilung mit Wahrscheinlichkeit mindestens $1-\delta$ jedes unbekannte Konzept $c \in \textsc{Intervall}^+$ mit Fehlerwahrscheinlichkeit höchstens ε lernt. Ihr Algorithmus sollte $\textsc{Intervall}^+$ auch als Hypothesenklasse benutzen und eine konsistente Hypothese ausgeben. Wenn Ihr Algorithmus von einer Hypothese nicht „überzeugt“ ist, darf er (im Gegensatz zu PAC-Algorithmen) weitere Beispiele anfordern.

   i) Geben Sie eine ausführliche Beschreibung Ihres Algorithmus an.

   ii) Wie viele Beispiele reichen aus, um mit hoher Wahrscheinlichkeit nur kleine Fehler zu machen?

   iii) Wie viele Beispiele sind dafür notwendig?

   Geben Sie für ii) bzw. iii) eine möglichst gute obere bzw. untere Schranke an.

   Hinweis: Wie können Sie Occams Razor hier einsetzen? Im Gegensatz zur Vorlesung ist es in diesem Fall nicht sinnvoll, über die Länge der binären Namen zu argumentieren.

   Bei der *Validierung* einer Hypothese h wird eine Menge V von s^* zusätzlichen Beispielen angefordert, um den globalen Fehler $\mathrm{fehler}_{\mathrm{D}}(c,h)$ approximativ zu bestimmen, wobei c das unbekannte Zielkonzept ist. Ist der mittels V geschätzte Fehler klein, gilt also $|V \cap (c \oplus h)| \leq s^* \cdot \varepsilon^*$ für ein geeignetes ε^*, so wird die Hypothese akzeptiert, andernfalls wird sie verworfen. Die Wahrscheinlichkeit, dass eine Hypothese h mit $\mathrm{fehler}_{\mathrm{D}}(c,h) > \varepsilon$ die Validierung dennoch besteht oder dass eine Hypothese h mit $\mathrm{fehler}_{\mathrm{D}}(c,h) \leq \varepsilon/4$ verworfen wird, soll höchstens δ^* sein. Bestimmen Sie s^* mithilfe der Chernoff-Ungleichung.

Fazit: Es gibt Konzeptklassen, für die es zwar keine PAC-Algorithmen gibt, die aber dennoch erfolgreich gelernt werden können, wenn eine variable Anzahl von Beispielen angefordert werden darf.

d*) Sei

$$\textsc{Intervall}^\infty := \bigvee_{i=1}^{\infty} \textsc{Intervall}$$

die Konzeptklasse aller Vereinigungen abzählbar vieler Intervalle. Gibt es einen Lernalgorithmus, der für jede Verteilung mit Wahrscheinlichkeit mindestens $1-\delta$ jedes unbekannte Konzept $c \in \textsc{Intervall}^\infty$ mit Fehlerwahrscheinlichkeit höchstens ε lernt? Begründen Sie Ihre Antwort.