AGML, Sommersemester 2020 Vorlesung 3 (6 Seiten)

Mengensysteme

Anne Driemel Letzte Aktualisierung: 27. April 2020

In den letzten Vorlesungen haben wir uns mit PAC-Lernbarkeit unter Annahme fester Hypothesenklassen beschäftigt. Wir haben gesehen, dass die Struktur einer Hypothesenklasse auch etwas über die Lernbarkeit aussagt, sofern die Hypothesenklasse realisierbar ist. In dieser Vorlesung werden wir uns mit der Struktur der Hypothesenklassen aus der Sicht von Mengensystemen befassen und allgemeine Eigenschaften ableiten. Wir nehmen dabei noch stets die Realisierbarkeit der Hypothesenklasse an.

1 Mengensysteme

Definition 3.1 (Mengensystem)**.** *Sei $\mathcal{X}$ eine beliebige Menge und $\mathcal{R}$ eine Menge von Teilmengen von $\mathcal{X}$. Wir nennen $\mathcal{R}$ ein Mengensystem mit Grundmenge $\mathcal{X}$.*

Jede Hypothesenklasse $\mathcal{H}$, definiert durch eine Menge von Funktionen der Form

$$h\colon \mathcal{X} \to \{-1, +1\},$$

kann gleichsam durch ein Mengensystem beschrieben werden. Wir definieren für jede Funktion $h \in \mathcal{H}$ eine Menge

$$r_h = \{\ x \in \mathcal{X}\ \mid\ h(x) = 1\ \}\,,$$

welche also genau der positiven Menge entspricht. Die Menge aller Mengen r_h bildet dann das Mengensystem.

Beispiel 3.2. *Die Menge aller achsenparallelen Rechtecke in der Ebene definiert ein Mengensystem $\mathcal{R}$ mit Grundmenge $\mathcal{X} = \mathbb{R}^2$. Formal ist jedes Element $r \in \mathcal{R}$ definiert durch ein Tupel (a, b, c, d) mit*

$$r_{a,b,c,d} = \{\ (x, y) \in \mathcal{X}\ \mid\ a \le x \le b, c \le y \le d\ \}\,.$$

Eine wichtige kombinatorische Eigenschaft von Mengensystemen ist ihre VC-dimension, benannt nach Vapnik und Chervonenkis.

Definition 3.3 (Abspalten)**.** *Wir sagen eine Menge $A' \subseteq \mathcal{X}$ wird durch ein Mengensystem $\mathcal{R}$ von einer Menge $A \subseteq \mathcal{X}$ abgespalten, wenn A' durch den Schnitt mit einer Menge von $\mathcal{R}$ erzeugt werden kann. Das heißt, es existiert ein $r \in \mathcal{R}$ mit $A' = r \cap A$.*

Definition 3.4 (Aufspalten)**.** *Eine Menge $A \subseteq \mathcal{R}$ wird durch ein Mengensystem aufgespalten, wenn alle Teilmengen von A abgespalten werden können.*

Definition 3.5 (VC-dimension)**.** *Die VC-dimension von $\mathcal{R}$ ist die Anzahl der Elemente in der größten durch $\mathcal{R}$ aufgespaltenen Menge. Falls keine solche Menge existiert, dann ist die VC-dimension unendlich. Wir bezeichnen die VC-dimension mit* $\dim(\mathcal{R})$*. Für den Sonderfall $\mathcal{R} = \emptyset$ definieren wir* $\dim(\emptyset) = 0$*.*

Schauen wir uns die VC-dimension im oben genannten Beispiel genauer an. Die VC-dimension von $\mathcal{R}$ ist mindestens 4, da wir eine 4-elementige Menge A von Punkten in der Ebene angeben können, die von $\mathcal{R}$ aufgespalten wird. Abbildung 1 zeigt eine solche Menge. Gleichzeitig können wir zeigen, dass für jede 5-elementige Menge A' gilt, dass sie *nicht* durch $\mathcal{R}$ aufgespalten wird.

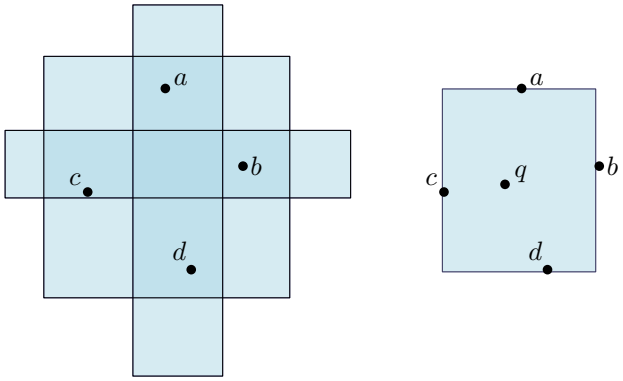

Abbildung 1: (links) Beispielmenge die durch das Mengensystem der achsenparallelen Rechtecke aufgespalten wird. Exemplarisch dargestellt sind auch drei achsenparallele Rechtecke, die verschiedene Teilmengen abspalten. (rechts) Bei fünf Punkten können wir immer einen Punkt q finden, sodass das Komplement nicht durch ein achsenparalleles Rechteck abgespalten werden kann.

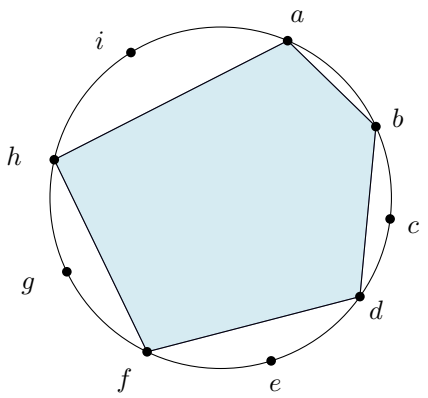

Abbildung 2: Beispielmenge von Punkten die durch das Mengensystem aller konvexer Polygone aufgespalten wird. Exemplarisch dargestellt ein Polygon, das die Teilmenge $\{a, b, d, f, h\}$ abspaltet.

Angenommen alle Koordinaten der Punkte in A sind paarweise verschieden. Da die Menge 5 Punkte enthält, existiert ein Punkt $q \in A$, der weder die x-Koordinate, noch die y-Koordinate in A minimiert oder maximiert. Es folgt, dass q in allen achsenparallelen Rechtecken enthalten ist, welche die Menge $A \setminus \{q\}$ enthalten. Daher existiert keine Menge $r \in \mathcal{R}$, sodass

$$A \setminus \{q\} = A \cap r.$$

Damit ist die VC-dimension des Mengensystems der achsenparallelen Rechtecke genau 4.

Es gibt auch Mengensysteme mit unendlicher VC-dimension. Betrachten wir das Mengensystem aller konvexen Polygone in der Ebene. Konvexe Polygone sind dadurch definiert, dass jeder Innenwinkel höchstens 180° beträgt. Für jede natürliche Zahl n können wir eine n-elementige Menge finden, welche durch dieses Mengensystem aufgespalten wird. Sei A_n eine Menge von n Punkten auf dem Einheitskreis. Jede Teilmenge $A \subseteq A_n$ definiert als Menge von Ecken ein konvexes Polygon P mit der gewünschten Eigenschaft, siehe Abbildung 2. Somit ist die VC-dimension des Mengensystems der konvexen Polygone unendlich.

2 Wachstum von endlichen Mengensystemen

Ein Mengensystem ist endlich, wenn es nur endlich viele Mengen enthält. Wieviele Mengen kann ein Mengensystem mit $\left|\mathcal{X}\right| = m$ enthalten? Allgemein gilt $\left|\mathcal{R}\right| \leq 2^{\left|\mathcal{X}\right|} = 2^m$. Was, wenn die VC-dimension kleiner als m ist?

Beispiel 3.6. *Sei $\mathcal{X} = \{1, 2, \ldots, m\}$ und sei $\mathcal{R}$ das Mengensystem, das alle Teilmengen von maximaler Größe k enthält. Die VC-dimension dieses Mengensystems ist k. Wir können alle generierten Mengen aufzählen und sehen direkt, dass*

$$\left|\mathcal{R}\right| = \sum_{i=0}^{k} \binom{m}{i} \leq \sum_{i=0}^{k} m^k \leq km^k.$$

Für ein festes k wächst die Anzahl der Mengen im Beispiel polynomiell in der Größe des Mengensystems m.

Wir wollen nun eine asymptotische obere Schranke zeigen, die dieses Wachstum im allgemeineren Fall von endlichen Mengensystemen mit endlicher VC-dimension beschreibt. Das folgende Lemma zeigt, dass die VC-dimension das Wachstum in diesem Sinne charakterisiert.

Lemma 3.7. *Es gilt für jedes Mengensystem $\mathcal{R}$ mit m-elementiger Grundmenge $\mathcal{X}$ und VC-dimension* d*, dass*

$$\left|\mathcal{R}\right| \leq \sum_{i=0}^{\mathrm{d}} \binom{m}{i}.$$

Beweis. Wir zeigen den Satz durch Induktion über m mit Induktionsanfang $m = 0$. In diesem Fall kann $\mathcal{R}$ höchstens die leere Menge enthalten, also ist $\left|\mathcal{R}\right| \leq 1$ und $\mathrm{d} \leq 0$. Gleichzeitig gilt per Definition des Binomialkoeffizienten, dass $\binom{0}{0} = 1$. Damit ist die Aussage für den Induktionsanfang erfüllt. Im Induktionsschritt nehmen wir an, dass $m > 0$. Sei $\mathcal{R}$ ein Mengensystem mit Grundmenge $\mathcal{X}$ und VC-dimension d. Nehmen wir an, dass $\mathrm{d} = 0$. In diesem Fall kann man auch zeigen, dass $\left|\mathcal{R}\right| \leq 1$ und die Aussage ist erfüllt. Also nehmen wir an, dass $\mathrm{d} > 0$.

Sei $x \in \mathcal{X}$ fest und betrachte das Mengensystem

$$\mathcal{R}_1 = \{\ r \setminus \{x\}\ \mid\ r \in \mathcal{R}\ \}.$$

Sei die VC-dimension d_1. Beachte, dass $\mathrm{d}_1 \leq \mathrm{d}$ ist, da jede Menge $A \subseteq \mathcal{X} \setminus \{x\}$ die durch $\mathcal{R}_1$ aufgespalten wird, auch durch $\mathcal{R}$ aufgespalten wird.

Nun folgt aus der Induktionsannahme, dass

$$\left|\mathcal{R}_1\right| \leq \sum_{i=0}^{\mathrm{d}_1} \binom{m-1}{i} \leq \sum_{i=0}^{\mathrm{d}} \binom{m-1}{i}.$$

Allerdings könnte es sein, dass zwei verschiedene Mengen in $\mathcal{R}$ durch die Beschränkung auf $\mathcal{X} \setminus \{x\}$ identisch werden und dadurch $\left|\mathcal{R}_1\right|$ strikt kleiner ist als $\left|\mathcal{R}\right|$. Wir definieren ein zweites Mengensystem um genau diese Paare von Mengen zu zählen, wie folgt

$$\mathcal{R}_2 = \{\ r \setminus \{x\}\ \mid\ r \setminus \{x\} \in \mathcal{R} \text{ und } r \cup \{x\} \in \mathcal{R}\ \}.$$

Es folgt nun, dass

$$\left|\mathcal{R}\right| = \left|\mathcal{R}_1\right| + \left|\mathcal{R}_2\right|.$$

Sei $\mathrm{d}_2 = \dim(\mathcal{R}_2)$. Wir behaupten, dass $\mathrm{d}_2 \leq \mathrm{d}-1$. Angenommen, dem wäre nicht so und die VC-dimension wäre mindestens d. Dann existierte eine Menge $A \subseteq \mathcal{X} \setminus \{x\}$ mit $|A| = \mathrm{d}$, sodass A durch $\mathcal{R}_2$ aufgespalten wird. Dann würde auch die Menge $A \cup \{x\}$ durch $\mathcal{R}$ aufgespalten, denn $\mathcal{R}_2$ enthält ja nur solche Paare von Mengen aus $\mathcal{R}$, die bis auf x identisch sind. Das würde aber der Grundannahme widersprechen, dass die VC-dimension von $\mathcal{R}$ gleich d ist.

Somit gilt nach Induktionsannahme, dass

$$|\mathcal{R}_2| \leq \sum_{i=0}^{\mathrm{d}_2} \binom{m-1}{i} \leq \sum_{i=0}^{\mathrm{d}-1} \binom{m-1}{i} = \sum_{j=1}^{\mathrm{d}} \binom{m-1}{j-1}$$

Durch Einsetzen in die obige Gleichung bekommen wir

$$|\mathcal{R}| \leq \sum_{i=0}^{\mathrm{d}} \binom{m-1}{i} + \sum_{j=1}^{\mathrm{d}} \binom{m-1}{j-1} = 1 + \sum_{i=1}^{\mathrm{d}} \binom{m-1}{i} + \binom{m-1}{i-1} = \sum_{i=0}^{\mathrm{d}} \binom{m}{i},$$

wobei die letzte Gleichung aus der rekursiven Darstellung des Binomialkoeffizienten folgt.

□

3 Unendliche Mengensysteme

Wir wollen nun die obige Schranke erweiteren auf unendliche Mengensysteme. Also Mengensysteme, die unendlich viele Mengen enthalten. Insbesondere sind wir interessiert am Wachstum der Anzahl der durch das Mengensystem abgespaltenen Teilmengen. Um das zu formalisieren betrachten wir Untersysteme, die wir wie folgt definieren.

Definition 3.8 (Untersystem)**.** *Sei $\mathcal{R}$ ein Mengensystem mit Grundmenge $\mathcal{X}$. Jede Menge $A \subseteq \mathcal{X}$ bestimmt ein Untersystem von $\mathcal{R}$ wie folgt*

$$\mathcal{R}|_A = \{\ r \cap A \ \mid \ r \in \mathcal{R}\ \}.$$

Das heißt, $\mathcal{R}|_A$ ist ein Mengensystem mit Grundmenge A, welches genau die Teilmengen von A enthält, die von A durch $\mathcal{R}$ abgespalten werden können. Die VC-dimension kann durch die Beschränkung auf ein Untersystem nicht größer werden, also gilt $\dim(\mathcal{R}|_A) \leq \dim(\mathcal{R})$.

Beispiel 3.9. *Wir haben Untersysteme schon kennengelernt, auch wenn wir sie nicht so genannt haben. Insbesondere ist das Mengensystem $\mathcal{R}_1$ aus vorhergehendem Beweis das Untersystem von $\mathcal{R}$ beschränkt auf $\mathcal{X} \setminus \{x\}$, denn,*

$$\mathcal{R}_1 = \{\ r \setminus \{x\} \ \mid \ r \in \mathcal{R}\ \} = \{\ r \cap (\mathcal{X} \setminus \{x\}) \ \mid \ r \in \mathcal{R}\ \} = \mathcal{R}|_{\mathcal{X}\setminus\{x\}}.$$

Satz 3.10 (Wachstumslemma)**.** *Sei $\mathcal{R}$ ein Mengensystem mit Grundmenge $\mathcal{X}$ und VC-dimension* d*. Für jede natürliche Zahl m gilt, dass*

$$\Pi_{\mathcal{R}}(m) = \max_{\substack{A \subseteq \mathcal{X} \\ |A| = m}} \left|\mathcal{R}|_A\right| \leq \left(\frac{\mathrm{e}m}{\mathrm{d}}\right)^{\mathrm{d}}.$$

Wir nennen $\Pi_{\mathcal{R}}$ die Wachstumsfunktion von $\mathcal{R}$.

Beweis. Da die VC-dimension durch die Beschränkung auf ein Untersystem nicht größer werden kann, können wir Lemma 3.7 direkt anwenden und bekommen für jede Menge $A' \subseteq \mathcal{X}$ mit $\left|A'\right| = m$, dass

$$\left|\mathcal{R}|_{A'}\right| \leq \sum_{i=0}^{\mathrm{d}} \binom{m}{i}.$$

Nun machen wir folgende Abschätzung

$$\binom{m}{i} = \frac{m!}{(m-i)! \cdot i!} \leq \frac{m^i}{i!} = \left(\frac{m}{\mathrm{d}}\right)^i \frac{\mathrm{d}^i}{i!} \leq \left(\frac{m}{\mathrm{d}}\right)^{\mathrm{d}} \frac{\mathrm{d}^i}{i!} \ .$$

Zusammen mit der Reihendefinition der Exponentialfunktion $\mathrm{e}^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$, bekommen wir dann

$$\sum_{i=0}^{\mathrm{d}} \binom{m}{i} \leq \sum_{i=0}^{\mathrm{d}} \left(\frac{m}{\mathrm{d}}\right)^{\mathrm{d}} \frac{\mathrm{d}^i}{i!} = \left(\frac{m}{\mathrm{d}}\right)^{\mathrm{d}} \sum_{i=0}^{\mathrm{d}} \frac{\mathrm{d}^i}{i!} \leq \left(\frac{m}{\mathrm{d}}\right)^{\mathrm{d}} \mathrm{e}^{\mathrm{d}} \ .$$

Da wir die Schranke für jede m-elementige Menge A' zeigen, gilt sie auch für die größte solche Menge. Damit ist der Satz bewiesen. □

4 PAC-Lernbarkeit

Unsere Motivation um Mengensysteme zu studieren war zu Beginn mit der Hoffnung auf bessere Schranken für die PAC-Lernbarkeit von Hypothesenklassen begründet. Sei $\mathcal{H}$ eine Hypothesenklasse und sei $\mathcal{R}$ das entsprechende Mengensystem mit Grundmenge $\mathcal{X}$. Sei $S \subseteq \mathcal{X}$. Schauen wir uns die Definitionen von $\mathcal{R}|_S$ und $\mathcal{H}|_S$ genauer an, sehen wir, dass diese äquivalent im Sinne unserer Abbildung zwischen Hypothesenklassen und Mengensystemen sind. Insbesondere beschreibt die Funktionsmenge $\mathcal{H}|_S$ alle verschiedenen Wege, Labels in $\{-1, +1\}$ für die Menge S zu vergeben mithilfe einer Funktion in $\mathcal{H}$. Im Kontext von Mengensystemen entspricht $\mathcal{R}|_S$ alle verschiedenen Wege, mithilfe einer Menge $r \in \mathcal{R}$ eine Teilmenge von S abzuspalten. Diese Teilmengen entsprechen dann den positiven Teilmengen von S, die sich aus Funktionen in $\mathcal{H}$ ergeben. Somit gilt für die Wachstumsfunktion

$$\Pi_{\mathcal{H}}(m) = \max_{S \subseteq X, |S|=m} \left|\mathcal{H}|_S\right| = \max_{S \subseteq X, |S|=m} \left|\mathcal{R}|_S\right| = \Pi_{\mathcal{R}}(m) \tag{1}$$

Satz 3.11. *Sei $\mathcal{H}$ eine Hypothesenklasse mit VC-dimension* d. *Seien $1 \geq \epsilon > 0$ und $\delta > 0$ beliebig und sei*

$$m \geq \max\left(\frac{4}{\epsilon}\log_2\frac{2}{\delta}, \frac{8\mathrm{d}}{\epsilon}\log_2\frac{16}{\epsilon}\right) \tag{2}$$

Betrachte ein Sample S von m Datenpunkten mit korrekten Labels gemäß f gezogen unabhängig und identisch verteilt aus $\mathcal{D}$. Es gilt mit Wahrscheinlichkeit mindestens $1 - \delta$, dass alle $h \in \mathcal{H}$ mit $\mathrm{err}_S(h) = 0$ auch $\mathrm{err}_{\mathcal{D},f}(h) < \epsilon$ erfüllen.

Beweis. Wir wollen Satz 2.7 aus der letzten Vorlesung anwenden, der besagt, dass die obige Behauptung gilt, sofern die folgende Bedingung für m erfüllt ist.

$$m \geq \max\left\{\frac{8}{\epsilon}, \frac{2}{\epsilon}\log_2\left(\frac{2\Pi_{\mathcal{H}}(2m)}{\delta}\right)\right\} \ . \tag{3}$$

Dafür müssen wir nur zeigen, dass unsere Bedingung (2) die Bedingung (3) impliziert.

Zunächst haben wir, da $\mathrm{d} \geq 1$ und $\epsilon \leq 1$,

$$m \geq \frac{8\mathrm{d}}{\epsilon}\log_2\frac{16}{\epsilon} \implies m \geq \frac{8}{\epsilon}$$

womit der erste Teil von Bedingung (3) gezeigt ist.

Wir behaupten nun, dass aus Bedingung (2) auch folgt, dass

$$m \geq \frac{2}{\epsilon} \log_2 \left(\frac{2}{\delta} \cdot \left(\frac{2\mathrm{e}m}{\mathrm{d}} \right)^{\mathrm{d}} \right). \tag{4}$$

Aus der Gleichheit der Wachstumsfunktionen von Hypothesenklassen und den zugehörigen Mengensystemen in (1) und der Schranke aus dem Wachstumslemma (Satz 3.10) folgt

$$\left(\frac{2\mathrm{e}m}{\mathrm{d}} \right)^{\mathrm{d}} \geq \Pi_{\mathcal{H}}(2m)$$

und somit wäre

$$m \geq \frac{2}{\epsilon} \log_2 \left(\frac{2\Pi_{\mathcal{H}}(2m)}{\delta} \right)$$

Damit wäre auch der zweite Teil der Bedingung 3 gezeigt.

Es bleibt, die Behauptung (2) $\implies$ (4) zu zeigen. Dafür formen wir (4) zunächst wie folgt um.

$$m \geq \frac{2}{\epsilon} \log_2 \frac{2}{\delta} + \frac{2}{\epsilon} \log_2 \left(\left(\frac{2\mathrm{e}m}{\mathrm{d}} \right)^{\mathrm{d}} \right). \tag{5}$$

Zunächst haben wir für die Abschätzung des ersten Terms

$$m \geq \frac{4}{\epsilon} \log_2 \frac{2}{\delta} \implies \frac{m}{2} \geq \frac{2}{\epsilon} \log_2 \frac{2}{\delta} \tag{6}$$

Für die Abschätzung des zweiten Terms zeigen wir

$$\frac{m}{2} \geq \frac{2}{\epsilon} \mathrm{d} \log_2 \left(\frac{2\mathrm{e}m}{\mathrm{d}} \right) \tag{7}$$

Setzen wir zunächst $m = \frac{8\mathrm{d}}{\epsilon} \log_2 \frac{16}{\epsilon}$ auf beiden Seiten ein, sehen wir durch äquivalente Umformung, dass die Ungleichung für $0 < \epsilon \leq 1$ erfüllt ist. Das gilt auch für größere Werte von m. Die Behauptung in (5) folgt nun durch das Addieren der beiden Ungleichungen in (6) und (7). □

Referenzen

- Foundations of Machine Learning, Kapitel 3.3
- Understanding Machine Learning, Kapitel 6.2-6.5 (anderer Beweis!)
- Sariel Har-Peled, Geometric Approximation Algorithms. AMS Mathematical Surveys and Monographs, Band 173. 2011.