10. Aufgabenblatt zur Vorlesung

# Randomisierte Algorithmen — WiSe 2016/17

Wolfgang Mulzer, Paul Seiferth, Max Willert

**Abgabe** bis zum 9. Januar 2017, 12 Uhr, in den Briefkasten neben Raum 111

**Aufgabe 1** Chans Optimierungstechnik — *10 Punkte*

(a) Sei $P \subset \mathbb{R}^2$ eine Menge von $n$ Punkten. Entwickeln und Analysieren Sie einen Algorithmus der den kleinsten Umfang eines Dreiecks berechnet, welches von drei Punkten aus $P$ aufgespannt wird. Ihr Algorithmus soll Chans Optimierungstechnik verwenden und in $O(n \log n)$ erwarteter Zeit laufen.

(b) Funktioniert Ihr Ansatz auch um die kleinste Fläche eines Dreiecks aus $P$ zu berechnen? Welche Schwierigkeiten treten auf?

**Aufgabe 2** Die Splitter-Funktion — *10 Punkte*

Sei $A = (X, \mathcal{R})$ ein Range Space, mit $X$ unendlich. Die *Splitterfunktion* für $A$, $\pi_A : \mathbb{N} \to \mathbb{N}$ ist definiert als

$$\pi_A(n) = \max_{Y \subset X, |Y|=n} |\mathcal{R}_Y|.$$

Beweisen Sie folgende Aussagen.

(a) Wenn $A$ VC-Dimension $d$ hat, dann ist $\pi_A(n) = O((en/d)^d)$.

(b) Falls ein $d \geq 0$ existiert, so dass $\pi_A(n) = O(n^d)$ ist, so hat $A$ die VC-Dimension $O(d \log d)$.

(c) Entweder existiert eine Konstante $d \geq 0$, so dass $\pi_A(n) = O(n^d)$ ist, oder es gilt $\pi_A(n) = 2^n$.

(d) Seien $A_1 = (X, \mathcal{R}_1)$ und $A_2 = (X, \mathcal{R}_2)$ Range Spaces. Wir definieren die Range Spaces $A_1 \cup A_2$, $A_1 \cap A_2$ und $A_1 \setminus A_2$ mit Grundmenge $X$ und Ranges $\{R_1 \cup R_2 \mid R_1 \in \mathcal{R}_1, R_2 \in \mathcal{R}_2\}$, $\{R_1 \cap R_2 \mid R_1 \in \mathcal{R}_1, R_2 \in \mathcal{R}_2\}$ bzw. $\{R_1 \setminus R_2 \mid R_1 \in \mathcal{R}_1, R_2 \in \mathcal{R}_2\}$. Zeigen Sie: Falls $A_1$ und $A_2$ beschränkte VC-Dimension haben, so auch $A_1 \cup A_2$, $A_1 \cap A_2$ und $A_1 \setminus A_2$.

**Aufgabe 3** PAC-Lernen — *10 Punkte*

Sei $X$ eine Menge. Eine Menge $\mathcal{C} \subseteq \{0,1\}^X$ von Funktionen von $X$ nach $\{0,1\}$ heißt *Konzeptklasse*. Eine Konzeptklasse heißt *PAC-lernbar*, falls ein *Lernalgorithmus* $A$ mit folgenden Eigenschaften existiert. Es existiert eine Funktion $k(\varepsilon, \delta) : (0,1]^2 \to \mathbb{N}$, so dass für alle $\varepsilon, \delta \in (0,1]$ und für alle Verteilungen $\mu$ auf $X$ gilt: Sei $f \in \mathcal{C}$ fest, aber unbekannt. Ziehe $k = k(\varepsilon, \delta)$ Samples $x_1, x_2, \ldots, x_k \in X$ gemäß der Verteilung

$\mu$. Bei Eingabe $(x_1, f(x_1)), (x_2, f(x_2)), \ldots, (x_k, f(x_k))$ liefert $A$ eine Funktion $g \in \mathcal{C}$, so dass mit Wahrscheinlichkeit mindestens $1 - \delta$ über die Wahl des Samples

$$\mathbf{Pr}_{x \sim \mu}[f(x) \neq g(x)] \leq \varepsilon$$

ist.

(a) Interpretieren Sie PAC-Lernbarkeit im Kontext des betreuten Lernens.

(b) Geben Sie eine geeignete Definition von VC-Dimension für Konzeptklassen.

(c) Zeigen Sie: Wenn eine Konzeptklasse beschränkte VC-Dimension hat, dann ist sie PAC-lernbar.

*Hinweis*: Verwenden Sie das verallgemeinerte $\varepsilon$-Netz Theorem und Aufgabe 2.

(d) Zeigen Sie: Sei $\mathcal{C}$ eine Konzeptklasse mit VC-Dimension mindestens $2z$. Dann existiert für jeden festen Lernalgorithmus $A$ eine Funktion $f \in \mathcal{C}$ und eine Verteilung $\mu$ auf $X$, so dass für mit Wahrscheinlichkeit mindestens $1/7$ für ein Sample $x_1, \ldots, x_z \sim \mu$ aus $X$

$$\mathbf{Pr}_{x \sim \mu}[f(x) \neq g(x)] \geq 1/8$$

ist, wobei $g$ die Ausgabe von $A$ ist.

*Hinweis*: Sei $Y \subseteq X$ eine zersplittere Teilmenge der Größe $2z$. Wählen Sie $\mu$ als die Gleichverteilung auf $Y$. Wählen Sie $f : Y \to \{0, 1\}$ als eine zufällige Funktion und zeigen Sie dann

$$\mathbf{E}_{f, x_1, \ldots, x_z \sim \mu}\left[\mathbf{Pr}_{x \sim \mu}[f(x) \neq g(x)]\right] \geq 1/4.$$

(e) Zeigen Sie: PAC-Lernbarkeit und beschränke VC-Dimension für Konzeptklassen sind äquivalent.