Hans U. Simon
Malte Darnstädt

Bochum, den 02.07.2013
Abgabe bis zum 10.07.2013

Übungen zur Vorlesung

Theorie des Maschinellen Lernens

Sommer 2013

Übungsblatt 11

Aufgabe 11.1

Wir stellen ein eingeschränktes Modell für die reelle Regression auf:
Sei X eine Menge, $B \geq 1$ und F eine Menge von Funktionen von X nach $[-B, B]$. Sei $t \in F$ das Zielkonzept und P eine Verteilung über $X \times [-B, B]$, so dass für alle P-verteilten Tupel x, y gilt: $t(x) = y$. Den Fehler einer Hypothese $f \in F$ definieren wir so wie in der Vorlesung.

Gib einen sinnvollen Lerner L in diesem Modell an und beweise für endliche F eine obere Schranke für m_L.

Aufgabe 11.2

Verallgemeinere die obere Schranke für totalbeschränkte Klassen aus der Vorlesung auf Verteilungen mit einem beliegen $B \geq 1$. Jetzt erlauben wir also Verteilungen P über $X \times [1 - B, B]$. Zeige:

Sei F eine bezüglich der L_∞-Metrik totalbeschränkte Klasse von Funktionen, die von X nach $[0, 1]$ abbilden, und sei $\mathcal{A}$ ein approximativer SEM-Algorithmus für F. Dann ist L mit $L(z) := \mathcal{A}(z, \epsilon_0/12)$ und $\epsilon_0 := B^2 \cdot \sqrt{288/m}$ ein Lerner für F mit Stichprobengröße

$$m_L(\epsilon, \delta, B) \leq \frac{288\, B^4}{\epsilon^2} \ln\left(\frac{2\,\mathcal{N}(\epsilon/(12B), F, d_{L_\infty})}{\delta}\right) \quad .$$

Gib nur die Änderungen zum Beweis aus der Vorlesung an.

Aufgabe 11.3

Zeige, dass die Voraussetzung *"die Funktionen in F sind beschränkt"* zum Lernen im Regressionsmodell notwendig ist. Es reicht, folgendes zu zeigen:

Es gibt ein F, das unbeschränkte Funktionen enthält, so dass man nicht mit einer polynomiellen Anzahl an Beispielen (polynomiell in $1/\epsilon$, $\log 1/\delta$ und $\log |F|$) lernen kann, selbst wenn $|F|$ endlich ist und wir nur Verteilungen P auf $X \times [0, 1]$ betrachten.

Aufgabe 11.4

Sei P eine Verteilung über $X \times \mathbb{R}$ und (x, y) sei P-verteilt. Zeige, dass der Minimierer von $\mathrm{er}_P(f)$ genau $f(x) = E[y|x]$ ist.

Hinweis: Es reicht zu zeigen, dass

$$E[(f(x) - y)^2] = E[(f(x) - E[y|x])^2] + E[(y - E[y|x])^2] \ .$$

Beachte die Regeln des bedingten Erwartungswertes für Zufallsvariablen A, B, C:

- $E[A|B]$ ist eine Funktion in B
  (insbesondere ist $E[A|B]$ selbst eine Zufallsvariable)
- Linearität: $E[\alpha \cdot A + \beta \cdot B|C] = \alpha \cdot E[A|C] + \beta \cdot E[B|C]$ mit $\alpha, \beta \in \mathbb{R}$
- Wenn A stochastisch unabhängig von B ist: $E[A|B] = E[A]$
- Satz der iterierten Erwartung: $E[E[A|B]] = E[A]$
- Wenn A eine Funktion in C ist (genauer: wenn A $\sigma(C)$-messbar ist): $E[A|C] = A$
  (zum Beispiel gilt $E[A|A] = A$; auch für Konstanten $\alpha \in \mathbb{R}$ folgt $E[\alpha|C] = \alpha$;
  für Produkte gilt $E[A \cdot B|C] = A \cdot E[B|C]$)