

Übungen zur Vorlesung
Theorie des maschinellen Lernens
Sommer 16
Übungsblatt 01

Aufgabe 1.1 (4 Punkte)

Sei $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ die Trainingsmenge. Die Hypothese

$$h_S(x) = \begin{cases} y_i & , \text{ falls } \exists i \in [m] : x_i = x \\ 0 & , \text{ sonst} \end{cases}$$

kann zu Overfitting führen. Sie sieht allerdings konstruiert aus.

Beachte: Wir betrachten Polynome in d Variablen!

a) Finde zu jedem $a \in \mathbb{R}^d$ ein Polynom p_a mit

$$p_a(x) \begin{cases} = 0 & , x = a \\ > 0 & , \text{ sonst} \end{cases} .$$

b) Zeige, dass $h_S(x)$ durch ein Schwellwertpolynom ersetzt werden kann, d.h.:

Sei $S = \{(x_1, y_1), \dots, (x_m, y_m)\} \subseteq (\mathbb{R}^d \times \{0, 1\})^m$ die gegebene Trainingsmenge. Gib ein Polynom p_S mit

$$h_S(x) = 1 \Leftrightarrow p_S(x) \geq 0$$

an.

Aufgabe 1.2 (4 Punkte)

Sei \mathcal{H} eine Klasse von Hypothesen der Form $h : X \rightarrow \{0, 1\}$. Seien D die zugrundeliegende (unbekannte) Verteilung und f die Zielfunktion. Sei $h \in \mathcal{H}$ fest. Zeige:

$$\mathbb{E}_{S|X \sim D^m} [L_S(h)] = L_{(D,f)}(h).$$

Welche Eigenschaft des Erwartungswertes nutzt du aus?

— Bitte wenden! —

Aufgabe 1.3 (4 Punkte)

Im Folgenden betrachten wir die Klasse der linearen Halbräume in \mathbb{R}^4

$$\mathcal{H} := \{h_{w,b} \mid w \in \mathbb{R}^4 \wedge b \in \mathbb{R}\},$$

wobei

$$h_{w,b}(x) := \begin{cases} 1 & , \text{ falls } \sum_{i=1}^4 w_i x_i \geq b \\ 0 & , \text{ sonst} \end{cases}.$$

Es steht folgende Menge an Trainingspunkten zur Verfügung

$$T := \{((4, 3, 4, 2), 1), ((4, 3, 1, 2), 1), ((1, 3, 4, 2), 0), ((2, 3, 4, 2), 1), \\ ((4, 3, 3, 2), 0), ((4, 3, 4, 1), 0)\}.$$

- a) Finde eine dreielementige Teilmenge von T mit folgenden Eigenschaften
- die Punkte haben nicht alle das gleiche Label,
 - die Punkte werden von einer Hypothese h aus \mathcal{H} korrekt gelabelt.

Gib die drei Punkte und die passende Hypothese h an.

- b) Finde eine dreielementige Teilmenge von T mit folgenden Eigenschaften
- die Punkte haben nicht alle das gleiche Label,
 - die Punkte werden von keiner Hypothese h aus \mathcal{H} korrekt gelabelt.

Gib die drei Punkte an und begründe warum es keine passende Hypothese gibt.

Aufgabe 1.4 (4 Punkte)

Sei $X = \{0, 1, \dots, n-1\}$. Wir betrachten die Hypothesenklasse

$$\mathcal{H}_{k,n} := \{h_a(x) \mid a \in \mathbb{R}^{k+1}\},$$

die durch das Vorzeichen von Polynomen (in einer Variablen) vom Grad k gebildet wird. Es gilt

$$h_a(x) := \begin{cases} 1 & , \text{ falls } \sum_{i=0}^k a_i x^i \geq 0 \\ 0 & , \text{ sonst} \end{cases}.$$

Wieviele verschiedene Hypothesen enthält $\mathcal{H}_{k,n}$ für X ?

Wieviele Beispiele reichen bei festem k für eine ERM-Hypothese h^* aus, damit mit hoher Wahrscheinlichkeit (mindestens $1 - \delta$) für eine beliebige Verteilung D und eine Zielfunktion $f \in \mathcal{H}$ gilt

$$L_{(D,f)}(h^*) \leq \varepsilon?$$

Informationen zu den Übungen

- Termine der Vorlesung:
 - Dienstag 12:00–14:00 in NA 1/64
 - Mittwoch 12:00–14:00 in NA 1/64
 - Termin der Übungsgruppe:
 - Dienstag 8:30–10:00 in NA 5/24
- Die Übungsgruppe wird von Christoph Ries geleitet. Seine Sprechstunde ist dienstags von 10 bis 11 Uhr in NA 1/69.
- Jeden Dienstag wird ein Übungsblatt auf der Internetseite

http://www.ruhr-uni-bochum.de/lmi/lehre/ml_ss16/

zur Verfügung gestellt.

- Die Abgabe der Lösungen und die Rückgabe der korrigierten Aufgaben erfolgt Dienstag morgens in der Übung. Falls eine Teilnahme an der Übung nicht möglich ist, können Aufgaben bei Christoph Ries in NA 1/69 eingereicht werden.
- Die Blätter können in Gruppen von bis zu maximal drei Personen bearbeitet und abgegeben werden. Jedes Gruppenmitglied muss aber in der Lage sein, in der Übung die Aufgaben an der Tafel vorzurechnen.
- Bitte notiert auf jedem Blatt Namen und Matrikelnummer.
- Auf einem Übungsblatt gibt es in der Regel vier Aufgaben mit jeweils vier erreichbaren Punkten. Für einen unbenoteten Teilnahmechein sind 50% der Übungspunkte zu erreichen. Außerdem ist eine regelmäßige Teilnahme an der Übung (inklusive der Präsentation einer eigenen Lösung) notwendig.

Viel Erfolg bei den Übungen!