
Statistische Lerntheorie SoSe 2012

Abt. Neuroinformatik

Dr. F. Schwenker

7. Aufgabenblatt (Abgabe am 26.06.2012 in der Vorlesung)

16. Aufgabe (10) :

Implementieren Sie den Lernalgorithmus L zum Lernen von Halbgeraden (am besten in `matlab`).

Es soll Generalisierungsfehler ϵ , Konfidenz $1 - \delta$ und Stichprobengröße M hierfür untersucht werden.

Hierzu müssen folgende Funktionen implementiert werden:

1. Ein Zielkonzept r_θ muss erzeugt werden. θ wird gemäß der Gleichverteilung aus dem Intervall $[-1, 1]$ gezogen.
2. Eine Stichprobe $s(M, r_\theta)$ mit M Datenpunkten wird erzeugt. Die Datenpunkte sollen nach der Standardnormalverteilung \mathcal{N} gewählt, und gemäß r_θ mit Labels $+1$ und 0 markiert werden.
3. Mit der Stichprobe $s(M, r_\theta)$ und dem Halbgeraden-Lernalgorithmus L wird eine Halbgeraden-Funktion r_λ berechnet.
4. Der Generalisierungsfehler ϵ muss ermittelt werden. Dazu werden 100000 Datenpunkt gemäß \mathcal{N} erzeugt, gemäß r_θ gelabelt und dann durch r_λ klassifiziert. Hieraus wird dann ϵ näherungsweise ermittelt.

Bestimmen Sie die Stichprobenumfänge M für $\epsilon = 0.1, 0.01$ und $\delta = 0.1, 0.01, 0.001$ und überprüfen Sie die Resultate experimentell. Hierzu müssen die Experimente wiederholt werden, wählen sie $100/\delta$ Wiederholungen.

Hinweis: Für $\epsilon = 0.01$ und $\delta = 0.001$ folgt der theoretische Wert $M = 691$.