Hans U. Simon
Christoph Ries

Bochum, den 11. Juli 2017
Abgabe bis zum 18.07.2017

Übungen zur Vorlesung

Theorie des maschinellen Lernens

Sommer 17

Übungsblatt 11

Aufgabe 11.1 (4 Punkte)
Gegeben sei die Trainingsmenge

$$S = \{((-2,7),1),((-1,2),3),((0,3),1),((1,1),3),((2,4),1),\\ ((3,5),3),((4,2),2),((5,6),1),((6,0),2),((7,1),2)\}$$

über der Grundmenge $\mathbb{R}^2$ und der Labelmenge $\{1,2,3\}$. Nutze den All-Pairs Ansatz und bestimme bezüglich der Decision Stumps über $\mathbb{R}^2$ so eine Vorhersagefunktion h.
Wie werden mit diesem h die Punkte $(1,5)$ und $(5,3)$ klassifiziert?

Aufgabe 11.2 (4 Punkte)
Die Hypothese $h_w(x)$ für die Multiklassen-Kategorisierung mit strukturierten Objekten soll effizient bestimmt werden. Zeige dazu wie man mit Hilfe der Tabelle $M(x|w)$ und Backtracking die beste Zeichenkette y^* ermitteln kann. Siehe dazu auch Ende des Abschnitts 17.3 im Skript.

Aufgabe 11.3 (4 Punkte)
Zeige, dass $\mathcal{O}(r \log r)$ Zeit ausreicht um den Rangnummernvektor $\overline{y}$ aus $y \in \mathbb{R}^r$ zubestimmen. Gib dazu einen geeigneten Algorithmus an. Siehe dazu auch Anfang des Abschnitts 17.4 im Skript.

— Bitte wenden! —

Aufgabe 11.4 (4 Punkte)

Betrachte den folgenden Perzeptron-Algorithmus für Multiklassen:

Eingabe Trainingsmenge $(x_1, y_1), \ldots, (x_m, y_m) \in \mathcal{X} \times \mathcal{Y}$
Label-sensitive Abbildung $\Psi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$

Initialisierung Setze $w^{(1)} = (0, \ldots, 0) \in \mathbb{R}^d$.

Hauptschleife Für $t = 1, 2, 3, \ldots$ mache folgendes:
Falls ein $i \in [m]$ und ein $y \neq y_i$ existieren mit $\langle w^{(t)}, \Psi(x_i, y_i) \rangle \leq \langle w^{(t)}, \Psi(x_i, y) \rangle$, dann setze

$$w^{(t+1)} = w^{(t)} + \Psi(x_i, y_i) - \Psi(x_i, y)$$

anderenfalls gib $w^{(t)}$ aus und stoppe.

Nimm an, dass ein w^* existiert, sodass für alle i und alle $y \neq y_i$ gilt $\langle w^*, \Psi(x_i, y_i) \rangle \geq \langle w^*, \Psi(x_i, y) \rangle + 1$. Es sei $R := \max_{i,y} ||\Psi(x_i, y_i) - \Psi(x_i, y)||$. Zeige, dass dann der obige Algorithmus nach höchstens $(R||w^*||)^2$ Iterationen der Hauptschleife anhält.