Übungen Sequenz- und Expressionsdatenanalyse

Sommersemester 2004

Dr. Ivo Grosse

IPK, Gatersleben
Institut für Informatik,
Universität Halle

Blatt 11

Aufgabe 11.1 (6 Punkte)

Es soll die Paritätsfunktion für dreistellige Binärzahlen $\vec{x} \in \{0,1\}^3$ mittels eines MLPs mit der Signum-Funktion als Aktivierungfunktion realisiert werden.

(a) Visualisieren Sie das Problem (das Sie natürlich auch als Zwei-Klassen-Problem auffassen können.). Überlegen Sie, wie Sie mit <u>einem</u> Modellneuron in den 3D-Raum eine Trennebene legen können, was analog zur Trenngeraden im Zweidimensionalen ist.

(b) Argumentieren Sie weiterhin, daß Sie mindestens eine verdeckte Schicht im MLP benötigen.

(c) Geben Sie die Paritätsfunktion als disjunktive Normalform an und realisieren Sie damit eine MLP zu deren Berechnung.

Aufgabe 11.2 (10 Punkte) **Freiwillige Zusatzaufgabe:**

Benutzen Sie die Datensätze seq_1 und seq_3 zum Trainieren von MDD Modellen mit Schwellen T und U, und berechnen Sie für gegebenes T und U die Fläche unter der ROC Kurve, $F(T,U)$, für die Datensätze seq_2 und seq_4. Schreiben Sie für $T = 5, 10, 15, ..., 100$ und $U = 50, 100, 150, 200, 250$ die Werte $F(T,U)$ in eine 20×5 Matrix. Welche MDD Modelle neigen zum Übertraining? Welches MDD Modell liefert die genaueste Klassifizierung von Donorstellen und Nicht-Donorstellen? Drucken Sie – für das optimale MDD Modell – die Entscheidungsbäume für seq_1 und seq_3. Sei K die Anzahl der Blätter des Entscheidungsbaums für Datensatz seq_1, und sei L die Anzahl der Blätter des Entscheidungsbaums für Datensatz seq_3. Benutzen Sie nun seq_1 zum Trainieren eines PWM Mischmodells mit K Klassen, benutzen Sie seq_3 zum Trainieren eines PWM Mischmodells mit L Klassen, und plotten Sie die ROC Kurve für seq_2 und seq_4. Vergleichen Sie diese ROC Kurve mit der ROC Kurve für das MDD Modell.

Vertauschen Sie die Datensätze seq_1 und seq_3 mit den Datensätzen seq_2 und seq_4, und wiederholen Sie die Analyse. Wie robust ist die 10×5 Matrix? Wie robust ist der Entscheidungsbaum des optimalen MDD Modells? Wie robust ist die ROC Kurve der zum optimalen MDD Modell gehörigigen PWM Mischmodelle?

Abgabetermin: 7. Juli (in der Übung)