

Maschinelles Lernen II

11. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade

Sommer 2011

Ausgabe am: 14.07.11

Aufgabe 1

Students t-Verteilung

Einer Ihrer Kommilitonen hat einen Klassifikator f gelernt und die Fehlerrate

$$\hat{R}(f) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}[f(\mathbf{x}_i) \neq y_i]$$

auf einer unabhängigen Testmenge $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_{20}, y_{20})\}$ bestimmt. Er erzählt Ihnen, dass er ein zweiseitiges Konfidenzintervall zum Konfidenzniveau $\alpha = 5\%$ unter der Normalverteilungsannahme berechnet hat. Welches Niveau hat es nach der „korrekten“ Statistik tatsächlich bestimmt?

Aufgabe 2

AUC

Evaluieren Sie das lineare Model

$$\mathbf{w} = \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}$$

auf den folgenden Testdaten

ID	1	2	3	4	5
\mathbf{x}_i	$\begin{pmatrix} 0.02 \\ 0.99 \end{pmatrix}$	$\begin{pmatrix} 0.35 \\ 0.94 \end{pmatrix}$	$\begin{pmatrix} 0.60 \\ 0.80 \end{pmatrix}$	$\begin{pmatrix} 0.86 \\ 0.50 \end{pmatrix}$	$\begin{pmatrix} 0.99 \\ 0.09 \end{pmatrix}$
y_i	-1	+1	-1	+1	+1

Geben Sie für $f_{\mathbf{w}}$ die ROC-Kurve an und bestimmen Sie den AUC-Wert.