

Knowledge Discovery in Databases
 SS 2011

Übungsblatt 3: Bewertung von Klassifikatoren

Aufgabe 3-1 *Bewertung von Klassifikatoren*

Gegeben sei ein Datensatz mit bekannter Klassenzugehörigkeit der Objekte. Um die Qualität eines Klassifikators K zu ermitteln wurden die Objekte mittels K klassifiziert. Die Klassifikationsergebnisse sind in der folgenden Tabelle dargestellt.

ID	Objektklasse	$K(o)$
O_1	A	A
O_2	B	A
O_3	A	C
O_4	C	C
O_5	C	B
O_6	B	B
O_7	A	A
O_8	A	A
O_9	A	A
O_{10}	B	C
O_{11}	B	A
O_{12}	C	A
O_{13}	C	C
O_{14}	C	C
O_{15}	B	B

- Berechnen Sie anhand der obenstehenden Ergebnisse Precision und Recall jeder Klasse.
- Um ein vollständiges Maß für die Güte der Klassifikation bezüglich einer Klasse zu haben, wird häufig auch das sogenannte F_1 -Measure (harmonisches Mittel zwischen Precision und Recall) verwendet. Das F_1 -Measure für Klasse i ist wie folgt definiert:

$$F_1(K, i) = \frac{2 \cdot \text{Recall}(K, i) \cdot \text{Precision}(K, i)}{\text{Recall}(K, i) + \text{Precision}(K, i)}$$

Berechnen Sie das F_1 -Measure für alle Klassen.

- Da das F_1 -Measure nur klassenweise definiert ist, ist es nicht direkt dazu geeignet einen Überblick über die Gesamtgüte des Klassifikators zu geben. Daher mittelt man es häufig über alle Klassen auf. Dabei wird zwischen den folgenden 2 Methoden unterschieden:
 - Micro Average F_1 -Measure: Hierbei werden die Werte für TP , FP und FN über alle Klassen aufsummiert und anschließend werden Precision, Recall und F_1 -Measure gebildet.

- Macro Average F_1 -Measure: Precision and Recall werden für jede Klasse gebildet und anschließend wird der Durchschnitt über alle Klassen gebildet. Danach kann dann das F_1 -Measure gebildet werden.

Bilden Sie im obigen Beispiel das Micro Average und Macro Average F_1 -Measure. Was beobachten Sie?

Aufgabe 3-2 *Bewertung von Klassifikatoren*

Gegeben ein Datensatz D mit Objekten aus zwei Klassen A und B ($D = A \cup B$), die *völlig zufällig* erzeugt wurden. Zudem gibt es in diesem Datensatz für beide Klassen jeweils die gleiche Anzahl an Objekten, d.h. $|A| = |B|$. Der beste Klassifikator kann daher immer nur die Klasse mit den meisten Objekten vorhersagen.

- Welche echte *Fehlerrate* ist von so einem *optimalen* Klassifikator zu erwarten?
- Welche Fehlerraten sind bei der Evaluation dieses Klassifikators beim Leave-one-out Test und dem 0.632 Bootstrap Verfahren zu erwarten? Interpretieren Sie die Resultate.