

Prof. Dr. Katharina Morik,
Prof. Dr. Claus Weihs,
Dr. Wouter Duivesteijn,
M.Sc. Sarah Schnackenberg,
B.Sc. Melanie Dagge

Dortmund, 02.07.14
Abgabe: bis Do, 09.07.2015,
10 Uhr, an
wouter.duivesteijn@tu-dortmund.de
und/oder in den Briefkasten "Duivesteijn"
im OH12, R4.005

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2015
Blatt 12

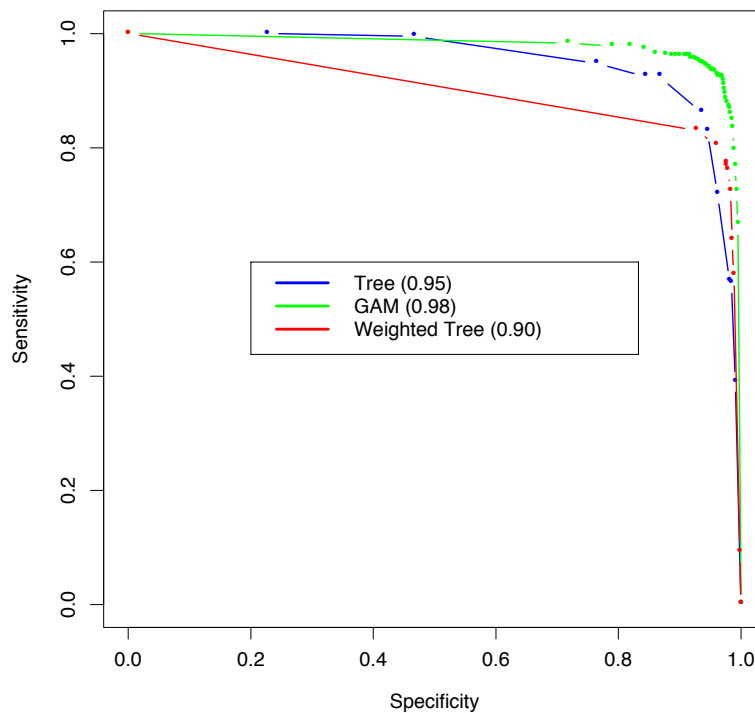


Abbildung 1: Example ROC space for Aufgabe 12.1

Aufgabe 12.1 (5 Punkte)

On June 30, we have learned about Subgroup Discovery.

1. (2 Punkte) What is the sensitivity of a rule? What is the specificity of a rule? How are these two quantities related in the ROC Curve?

For the remaining parts of this Aufgabe, consider the ROC space in Abbildung 1. Where in the Figure would we find the following rules:

2. (1 Punkt) the empty rule h_{\emptyset} , which is the rule covering no examples in the dataset;
3. (1 Punkt) the complete rule h_{Ω} , which is the rule covering all examples in the dataset;
4. (1 Punkt) the perfect rule h_p , which is the rule covering all positive and no negative examples in the dataset?

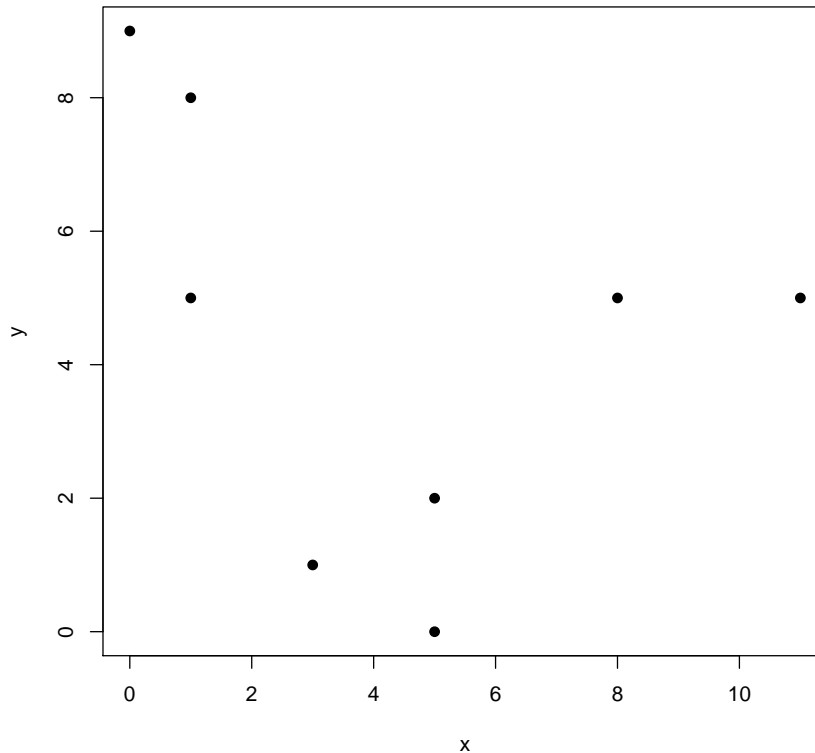


Abbildung 2: Example data for Aufgabe 12.2

Aufgabe 12.2 (5 Punkte)

Gegeben seien folgende Datenpunkte im Euklidischen Raum (siehe auch Abbildung 2):

Punkt	A	B	C	D	E	F	G	H
x	0	1	1	3	5	5	8	11
y	9	8	5	1	2	0	5	5

1. (2 Punkte) Führen Sie den K-Means-Algorithmus mit $k = 3$ per Hand aus. Normalerweise werden die Startpunkte für die Mittelpunkte der Cluster (auch Zentroiden genannt) zufällig gewählt. Hier sollen Sie jedoch die Punkte A, B und C als Startzentroiden benutzen. Falls im laufenden Algorithmus ein konkretes Beispiel äquidistant zu zwei Clusterzentroiden ist, so wählen Sie denjenigen, der näher am Nullpunkt liegt.

2. (1 Punkt) Benutzen Sie nun die Punkte E, F und G als Startzentroiden und führen Sie den Algorithmus ein weiteres Mal durch.
3. (2 Punkte) Berechnen Sie für Ihre Resultate aus a) und b) jeweils die in der Vorlesung auf Folie 17 vorgestellten Gütekriterien *Innerer Abstand* $W(C)$ und *Zwischenunähnlichkeit* $B(C)$. Interpretieren Sie kurz die Ergebnisse.