

Exercise „Business Analytics“ - SS 2012 – Information Systems and Machine Learning Lab (ISMLL) - Dr. Tomas Horvath, Osman Akcatepe

Exercise Sheet 2

Deadline: Thursday, 17.05.2011 at 16.00

1. Distinguish between noise and outliers. Be sure to consider the following questions.

- (a) Is noise ever interesting or desirable? Outliers?
- (b) Can noise objects be outliers?
- (c) Are noise objects always outliers?
- (d) Are outliers always noise objects?
- (e) Can noise make a typical value into an unusual one, or vice versa?

2. With n data points $D=\{x_i, y_i\}$, we want to find a model that produces output y when given input x : $h(x)=y$.

- (a) How can we find this model?
- (b) After finding this model, what can we do to avoid overfitting?

3. Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

Example: Age in years. Answer: Discrete, quantitative, ratio

- (a) Time in terms of AM or PM.
- (b) Brightness as measured by a light meter.
- (c) Brightness as measured by people's judgments.
- (d) Angles as measured in degrees between 0° and 360° .
- (e) Bronze, Silver, and Gold medals as awarded at the Olympics.
- (f) Height above sea level.
- (g) Number of patients in a hospital.

(h) ISBN numbers for books.

(i) Ability to pass light in terms of the following values: opaque, translucent, transparent.

(j) Military rank.

(k) Distance from the center of campus.

(l) Density of a substance in grams per cubic centimeter.

4. You are given a set of m objects that is divided into K groups, where the i th group is of size m_i . If the goal is to obtain a sample of size $n < m$, what is the difference between the following two sampling schemes? (Assume sampling with replacement.)

(a) We randomly select $n * m_i / m$ elements from each group.

(b) We randomly select n elements from the data set, without regard for the group to which an object belongs.