Übung „Business Analytics“ - SS 2012 - Wirtschaftsinformatik und Maschinelles Lernen (ISMLL) - Dr. Tomas Horvath, Osman Akcatepe

Übungsblatt 2

Abgabe bis: Donnerstag, 17.05.2012, um 16.00 Uhr

1.Unterscheiden Sie zwischen Noise und Outliers für die folgende Aussagen.

(a) Ist Noise je interessant oder wünschenswert? Outliers?

(b) Können Noise-objekte Outliers sein?

(c) Sind Noise-objekte immer outliers?

(d) Sind Outliers immer Noise-objekte?

(e) Kann ein Noise einen typischen Wert in ein ungewöhnliches einer machen, oder auch umgekehrt?

2. Mit ***n*** Datenpunkten $D=\{x_i, y_i\}$ möchten wir ein Modell zu finden, die ein Output ***y*** produziert, when man das Input ***x***: ***h(x)=y*** eingibt.

(a)Wie können wir dieses Modell finden?

(b)Nach dem Finden dieses Modells, was können wir machen das Overfitting zu vermeiden?

3. Klassifizieren Sie das folgende Attribut (attribute) als binär (binary), diskret (discrete), oder stetig (continuous). Klassifizieren Sie auch sie als qualitativ (nominell oder ordinal) oder quantitativ (Abstand oder Verhältnis). Einige Fälle dürfen mehr als eine Interpretation haben; damit deuten Sie kurz Ihr Argument hin, wenn Sie denken, dass es eine Zweideutigkeit geben darf.

Beispiel: Älter in Jahren. Antwort:Diskret, quantitativ, Verhältnis.

(a) Zeit in Begriffen von AM oder PM.

(b) Helligkeit als gemessen durch einen leichten Meter.

(c) Helligkeit als gemessen durch die Urteile von Leuten.

(d) Winkel als gemessen in Graden zwischen 0 und 360.

(e) Bronze, Silber, und Goldmedaillen als gewährt an den Olympischen Spielen.

(f) Höhe über Meeresspiegel.

(g) Zahl von Patienten in einem Krankenhaus betrachtet ist.

(h) ISBN Zahlen für Bücher.

(i) Militärischer Rang.

(j) Entfernung von der Mitte des Campus.

(k) Dichtheit einer Substanz in Grammen pro kubischen Zentimeter.

4. Sei eine Menge von ***m*** Objekten gegeben, die in ***K*** Gruppen geteilt sind, wo die ***i***-te Gruppe von Größe m_i ist. Was ist der Unterschied zwischen dem Folgenden zwei Samplingschemata, wenn das Ziel ein Sample der Größe **n<m** zu erhalten ist? (Nehmen Sie Sampling mit Ersetzung an).

(a) Wählen wir zufällig $n * m_i/m$ Elemente von jeder Gruppe aus.

(b) Wir wählen zufällig ***n*** Elemente vom Datensatz abgesehen von der Gruppe aus, zu der ein Objekt gehört.