

Knowledge Discovery in Databases
SS 2015

Übungsblatt 12: Apriori, Kernel und Regression

Aufgabe 12-1 Apriori-Algorithmus

Gegeben ist die Menge der Items $I = \{A, B, C, D, E, F, G, H, I, K, L, M\}$.

Weiterhin ist eine Menge von Transaktionen T laut folgender Tabelle gegeben:

Menge der Transaktionen T

Transaktions ID	gekaufte Items
1	B E G H
2	A B C E G H
3	A B C E F H
4	B C D E F G H L
5	A B E K H
6	B E F G H I K
7	A B D G H
8	A B D G
9	B D F G
10	C E F
11	A C E F H
12	A B E G

- Bestimmen Sie zum minimalen Support von 30% die häufig auftretenden Itemsets. Verwenden Sie dazu den Apriori-Algorithmus. Geben Sie insbesondere die Kandidatenmengen nach den Join-Schritten und nach den Prune-Schritten an, sowie die häufig auftretenden Itemsets mit ihrem jeweiligen Support.
- Bestimmen sie alle Assziationregeln, die man aus dem frequent Itemset $\{B E G H\}$ mit einer Konfidenz von mindestens 60% und dem Support 4 (Frequency 30%) bilden kann. Nutzen sie dabei die in der Vorlesung vorgestellte Monotonie aus.

Aufgabe 12-2 Kernel-Funktionen

Wie in der Vorlesung erklärt, zeichnet sich eine Kernel-Funktion (“Kernel”) durch positive (Semi-)Definitheit aus. Eine Matrix A ist positiv definit, falls ihre Eigenwerte nichtnegativ sind, oder alternativ formuliert, falls für all $x \in \mathbb{R}^d$ gilt: $x^\top \cdot A \cdot x \geq 0$

Zeigen Sie, dass folgende Funktionen Kernels sind, falls x und \hat{x} Vektoren im \mathbb{R}^d sind:

(a) $k_1(x, \hat{x}) = 1$

(b) $k_2(x, \hat{x}) = 3 \cdot x^\top \cdot \hat{x}$

(c) $k_3(x, \hat{x}) = 3 \cdot x^\top \cdot \hat{x} + 5$

Aufgabe 12-3 Lineare Separierbarkeit

Geben Sie für die folgenden Booleschen Funktionen an, ob das entsprechende Problem linear separierbar ist.

(a) $A \wedge B \wedge C$

(b) $A \vee B$

(c) $(A \vee B) \wedge (A \vee C)$

(d) $\neg A \wedge B$

Aufgabe 12-4 Perceptron

Zeigen Sie, wie das Peceptron-Modell (kein hidden layer!) verwendet werden kann, um über zwei Boolesche Variablen $x_1, x_2 \in \{0, 1\}$ die UND bzw. die ODER Funktion ($x_1 \wedge x_2$ bzw. $x_1 \vee x_2$) zu repräsentieren.

Aufgabe 12-5 Lineare Regression

Das Gehalt einer Person hängt von der Anzahl der Jahre ab, in denen die Person ihren Beruf ausgeübt hat. Um diesen Zusammenhang genauer zu untersuchen, kann man ein lineares Regressionsmodell lernen. Als Trainingsmenge stehen uns die Jahre an Berufserfahrung und die Gehälter folgender Personen zur Verfügung.

Erfahrung in Jahren	Gehalt in (1000\$)
3	30
8	57
9	64
13	72
3	36
6	43
11	59
21	90
1	20
16	83

- (a) Berechnen Sie eine Regressionsgerade, die dazu dienen soll, das voraussichtliche Gehalt auf Basis der Berufserfahrung abzuschätzen. Bestimmen Sie hierzu die Gerade, die den quadratischen Fehler minimiert.
- (b) Bestimmen Sie den quadratischen Fehler der berechneten Gerade, um abzuschätzen, wie gut die Regressionsgerade den Zusammenhang erklärt.
- (c) Berechnen Sie mit Hilfe Ihrer Regressionsgerade das voraussichtliche Gehalt für Personen mit den folgenden Jahren an Berufserfahrung:
Person A: 20
Person B: 8
Person C: 11