

Data Mining

Dauer: 20 Minuten

Vorbereitungsphase

Wie lange sollte man sich auf die Klausur vorbereiten?

ca. 4 Tage

Wie hast du dich vorbereitet (allein, in der Gruppe)?

allein

Welche Vorlesungen gingen der Prüfung voraus / waren hilfreich?

Data Mining

Welche Literatur/Skripte waren hilfreich?

- Skript zur Vorlesung
- Seiten zu Data Mining anderer Unis
- Internet allgemein zur Verfahrensrecherche

Gibt es allgemeine Tipps, die bei der Vorbereitung helfen könnten?

- Übersicht zu den drei großen Bereichen erstellen, also Klassifikation + Evaluierung, Clustering + Evaluierung und Assoziationsregeln; steht in Klammern, weil mehr Wert auf die ersten beiden Bereich gelegt wurde)
- Formel wissen ist gut, aber nicht unbedingt nötig (einiges lässt sich aber anhand der Formeln besser erklären, z.B. SSE war ihr sehr wichtig)

Wo lagen Deiner Meinung nach besondere Schwierigkeiten der Klausur?

Es werden Anwendungsfragen gestellt, wie z.B.: "Stellen sie sich vor, sie wollen E-Mails clustern. Wie können sie die Distanz zwischen diesen Dokumenten berechnen?" Hier wird auch Wissen aus anderen Fächern benötigt, wie z.B. Dokumentverarbeitung (darstellen der Dokumente als Vektoren mit Wörtern als Dimension - > dann berechnen der Distanz durch Vektoren)

Verlauf der Prüfung

Wie verlief die Prüfung?

- sehr locker

Wie reagierte der Prüfer, wenn Fragen nicht sofort beantwortet wurden?

- gibt große Hilfen und Beispiele um die Fragestellung genauer zu erklären

Dein Kommentar zur Benotung:

gut

Welche Fragen wurden konkret gestellt?

Was ist das Ziel von Clustering?

Was wird beim Clustern optimiert? (intra-Distanz minimieren / inter-Distanz maximieren)

Wie funktioniert K-Means?

Was optimiert K-Means? Was nicht? (intra: Ja / inter: Nein)

Welche Zielfunktion wird bei K-Means optimiert? (SSE)

Frage nach dem E-Mail Clustering (Siehe oben)!

Was ist das Ziel der Klassifikation?

Wie funktioniert Naive Bayes?

Wie kann man die Qualität von Klassifikationen ermitteln?