

Knowledge Discovery in Databases
 WS 2010/11

Übungsblatt 4: Bayes-Klassifikation, Nächste-Nachbarn-Klassifikation

Aufgabe 4-1 *Naive Bayes*

Die Ski-Saison ist eröffnet. Um zuverlässig zu entscheiden, wann Sie Skifahren gehen können und wann nicht, können Sie einen Klassifikator (z.B. Naive Bayes) benutzen. Der Klassifikator wird mit Ihren Erfahrungswerten aus dem letzten Jahr trainiert. Berücksichtigt werden dabei folgende Attribute:

Das Wetter: Das Attribut `Wetter` kann die folgenden drei Werte annehmen: `Sonne`, `Regen` und `Schneefall`.

Die Schneehöhe: Das Attribut `Schneehöhe` kann die folgenden zwei Werte annehmen: ≥ 50 (Es liegen mindestens 50 cm Schnee) und < 50 (Es liegen weniger als 50 cm Schnee).

Angenommen, Sie wollten letztes Jahr 8-mal zum Skifahren gehen. Die folgende Tabelle gibt Ihre jeweiligen Entscheidungen wieder:

| Wetter | Schneehöhe | Skifahren ? |
|--------|------------|-------------|
| Sonne | < 50 | nein |
| Regen | < 50 | nein |
| Regen | ≥ 50 | nein |
| Schnee | ≥ 50 | ja |
| Schnee | < 50 | nein |
| Sonne | ≥ 50 | ja |
| Schnee | ≥ 50 | ja |
| Regen | < 50 | ja |

- Berechnen Sie die *a priori* Wahrscheinlichkeiten für die beiden Klassen `Skifahren = ja` und `Skifahren = nein` (auf den Trainingsdaten)!
- Berechnen Sie für die Klassen die Werteverteilungen aller Attribute.
- Entscheiden Sie, ob Sie bei den folgenden Wetter- und Schneebedingungen Skifahren gehen oder nicht! Verwenden Sie dazu den naiven Bayes-Klassifikator.

| | Wetter | Schneehöhe |
|-------|------------|------------|
| Tag A | Sonne | ≥ 50 |
| Tag B | Regen | < 50 |
| Tag C | Schneefall | < 50 |

Aufgabe 4-2 *Nächste-Nachbarn-Klassifikation*

Geben Sie eine Punktmenge an, bestehend aus mindestens vier 2-dimensionalen Punkten, so dass die Nächste-Nachbarn-Klassifikation ($k = 1$) auf diesen Punkten nur Fehlklassifikationen liefert! Als Distanzfunktion sei die euklidische Distanz gegeben.