

Vorlesung Machine Learning und Data Mining

Übungsblatt für den 12.1.2006

Aufgabe 1

Ein Patient weiß folgendes über einen bestimmten Krebstest: Falls jemand Krebs hat, ist der Test in 98% der Fälle korrekt. Falls jemand keinen Krebs hat, ist der Test in 97% der Fälle korrekt. Insgesamt haben 0,8% der gesamten Bevölkerung Krebs.

Der Patient erhält nun die Nachricht, daß sein Test positiv ist. Was sagt ihm das?

Gegeben sei folgende Beispielmenge:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	26	High		No
D2	Sunny	28	High	Strong	No
D3	Overcast	29	High	Weak	Yes
D4	Rain	23	High	Weak	Yes
D5	Rain		Normal	Weak	Yes
D6	Rain	12	Normal	Strong	No
D7	Overcast	8		Strong	Yes
D8	Sunny	25	High	Weak	No
D9	Sunny	18	Normal	Weak	Yes
D10	Rain	20	Normal	Weak	Yes
D11	Sunny	20	Normal	Strong	
D12	Overcast	21	High	Strong	Yes
D13		26	Normal	Weak	Yes
D14	Rain	24	High	Strong	No
D15	Sunny	23	Normal	Weak	No
D16	Sunny	21	Normal	Weak	Yes

Aufgabe 2

a) Berechnen Sie die Tabelle der bedingten Wahrscheinlichkeiten, wie sie Naïve Bayes erzeugt.

b) Welchen Klassifikationswert gibt Naïve Bayes für die folgenden Instanzen aus?

1. Outlook=Sunny, Temperature=23, Humidity=High, Wind=Strong
2. Day=D18, Outlook=Rain, Humidity=Normal
3. Temperature=28

Aufgabe 3

Betrachten sie folgende Regeln:

1. Outlook = Sunny \rightarrow Yes else No
2. Wind = Weak \rightarrow No else Yes
3. Humidity = Normal and $16 < \text{Temperature} < 25 \rightarrow$ Yes else No
4. Temperature $> 28 \rightarrow$ Yes else No

Ohne auf die Daten zu schauen, schätzen Sie bitte die Plausibilität jeder einzelnen Regel ein. Weisen Sie jeder Regel diesen Wert als Wahrscheinlichkeit zu.

Betrachten Sie nun den Datensatz:

- a) Welche der Regeln ist h_{MAP} , welche h_{ML} ?
- b) Wie lautet die Bayes'sche optimale Klassifikation für die Instanz Outlook = Sunny, Temperature=22, Humidity=High, Wind=Normal?

Aufgabe 4

- a) Überlegen Sie sich eine gute Abstandsfunktion für die einzelnen Attribute.
- b) Benutzen Sie 3-NN zum Ausfüllen der fehlenden Werte.

Beziehen Sie hier die Klassifikation mit ein oder nicht? Warum?

- b) Welchen Klassifikationswert gibt k -NN für die folgende Instanz aus?

1. Outlook=Sunny, Temperature=23, Humidity=High, Wind=Strong

Testen Sie verschiedene k . Für welches k ändert sich die Klassifikation gegenüber $k = 1$?

- c) Berechnen Sie den Klassifikationswert obiger Instanz mittels abstandsgewichtetem NN (Shepards Methode).