Maschinelles Lernen

3. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade
Jules Rasetaharison

WS10/11

Ausgabe am: 08.11.10
Besprechung am: 15.11.10

Aufgabe 1 (1/5 Punkt):
Erklären sie die Begriffe *Entropie* $H(y)$ und *Information Gain* $IG(x)$ an einem selbstgewählten Beispiel. Was bedeuten die Fälle $H(y) = 0$, $H(y) = 1$, $IG(x) = 0$ und $IG(x) = 1$? Als *bedingte Entropie* $H(y|x)$ bezeichnet man die Entropie der Zufallsvariable y gegeben x, wobei $H(x,y) = H(y|x) + H(x)$. Besteht ein Zusammenhang zwischen der bedingten Entropie und Information Gain?

Aufgabe 2 (1/5 Punkt):
Werfen sie einen Blick auf die folgende Trainingsmenge (y und z sind diskrete Attribute); simulieren sie C4.5 von Hand. Verwenden sie Information Gain (nicht Gain Ratio). Bevorzugen sie bei gleichem Information Gain diskrete Attribute vor kontinuierlichen; welchen Vorteil hat dies? Wie sieht die Berechnungsspur von C4.5 aus und welcher Baum wird gelernt?

x	y	z	c
0,1	2	0	A
0,1	3	0	A
0,2	3	1	A
0,2	4	0	B
0,3	2	0	A
0,35	2	1	B
0,3	3	1	B
0,3	4	1	B

Aufgabe 3 (1/5 Punkt):
Nehmen wir an, dass $T1$ ein Entscheidungsbaum ist, und $T2$ ein Entscheidungsbaum, der dadurch entsteht, dass statt eines Blattknotens ein weiterer Split-Knoten eingefügt wird. Besteht zwischen $T1$ und $T2$ eine "Allgemeiner-Als"-Beziehung (siehe erste Vorlesung)? (Falls ja, welche $T1 \geq_g T2$ oder $T2 \geq_g T1$) und unter welchen Annahmen? Begründen sie Ihre Antwort.

Aufgabe 4 (1/5 Punkt):
Nehmen wir an, ihr Onkel arbeitet bei einer Bank und sie möchten ihm helfen, die Kreditkrise durchzustehen. Er bittet sie, einen Entscheidungsbaum zu lernen und vorherzusagen, ob zukünftige Kreditantragsteller ihren Kredit zurückzahlen werden oder nicht. Die Trainingsdaten und die Testdaten finden sie auf der Vorlesungswebseite neben dem Übungsblatt im weka-Format. Installieren sie das weka-Toolkit (www.cs.waikato.ac.nz/ml/weka) und verwenden sie den J48-Algorithmus, dieser entspricht dem C4.5-Algorithmus aus der Vorlesung. Trainieren sie einen Entscheidungsbaum auf den Trainingsdaten mit den voreingestellten Standardparametern und klassifizieren sie die Testdaten. Wieviel Prozent der Testdaten werden falsch klassifiziert? Variieren sie den Pruningparameter "minNumObj"

bzw. "-M" im Intervall von 1-6 und notieren sie den jeweiligen Testfehler. Erklären sie den beobachteten Zusammenhang zwischen Pruningparameter und Testfehler.

Aufgabe 5 (1/5 Punkt):
In der Vorlesung wurde die Herleitung der mehrdimensionalen linearen Regression skizziert. Die ausführlichen Herleitungsschritte lauten folgendermassen.

$$SSE = \sum_j (y_j - \mathbf{w}^\top \mathbf{x}_j)^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}) = \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\mathbf{w} + \mathbf{w}^\top \mathbf{X}^\top \mathbf{X}\mathbf{w}$$

Ableiten:

$$\frac{\partial SSE}{\partial \mathbf{w}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\mathbf{w}$$

Nullstellen und auflösen nach $\mathbf{w}$:

$$0 = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\mathbf{w} \Leftrightarrow \mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

Wir fügen zu der obigen Zielfunktion noch einen Standard-Regularisierer $\lambda \mathbf{w}^\top \mathbf{w}$ hinzu und erhalten damit die Zielfunktion der so genannten *Ridge regression*:

$$SSE_{reg} = \sum_j (y_j - \mathbf{w}^\top \mathbf{x}_j)^2 + \lambda \mathbf{w}^\top \mathbf{w}$$

Leiten sie nun analog zu der Herleitung der unregularisierten linearen Regression einen Schätzer für den Parameter $\mathbf{w}$ der Ridge Regression her.