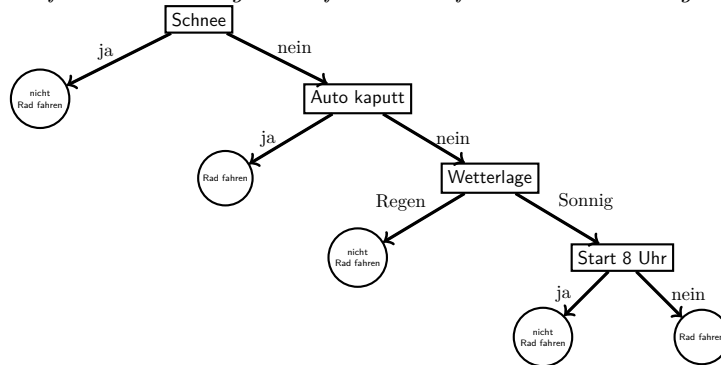


Übungsblatt 1

1. Baumeigenschaften



Tiefer Entscheidungsbaum für die Radfahren-Entscheidung



Berechnen Sie für diesen Baum, die **Anzahl der Blätter im Baum**, die **Baumhöhe**, die **External Path Length** und die **Weighted external path length**. Benutzen Sie für die **Weighted external path length** die Tabelle aus Seite 135 in Buch.

2. CART für den Fahrrad-Verleih



Nutzen Sie den bereits fertig implementierten CART auf einer etwas anderen Testmenge als in der Vorlesung.

Vorher entfernen Sie das Merkmal 3 (Day) und nutzen dies lediglich zur Aufteilung in Trainings- und Testset. Ebenfalls entfernen werden wir die Datensätze mit `Weathersit=4`, da es hierfür nur **3** Datensätze gibt. Diese führen nur zu Problemen. Für die Merkmale 0 (Jahreszeit) und 6 (Wochentag) können Sie ggf. dazu übergehen, es in eine nominale Kodierung zu überführen oder es in der vorhandenen Kodierung zu belassen.

Ziel: Unser Ziel ist es nun zu lernen, die Anzahl der Ausleihen insgesamt vorherzusagen. Hierzu werden wir immer die Daten vom 1. bis einschließlich zum 20. eines Monats zum Training verwenden und alle Tage ab dem 21. zum Testen.

Bitte wenden!

3. CART mit Block & Bleistift



Die Datensätze rechts entstammen dem „Two Moons Problem“.

1. Erstellen Sie einen Plot mit Python oder per Hand in dem diese 14 Punkte eingezeichnet sind. Der Plot sollte mindestens eine halbe DIN A4 Seite groß sein.
2. Führen Sie den CART-Algorithmus quasi per Hand durch. Rechnen Sie den ersten **best split** zu Fuß. Danach dürfen Sie aber die Methode **calGiniImpurity** sich heraus kopieren und zur Berechnung nutzen.
3. Tun Sie so, also wenn der Algorithmus mit den Werten **threshold = 0.1**, **xDecimals = 2**, **minLeafNodeSize=3** aufgerufen werden würde. Zeichnen Sie dabei die Schnitte in den Plot oben.
4. Zeichnen Sie die Gebiete die einer Klasse zugeordnet werden in den Plot ein.
5. Zeichnen Sie den resultierenden Baum auf

x_0	x_1	Klasse
1	0	0
0.9	0.5	0
0.5	0.9	0
0	1	0
-0.5	0.9	0
-0.8	0.5	0
-1	0	0
0	0.5	1
0.1	0	1
0.5	-0.4	1
1	-0.5	1
1.5	-0.4	1
1.9	0	1
2	0.5	1