

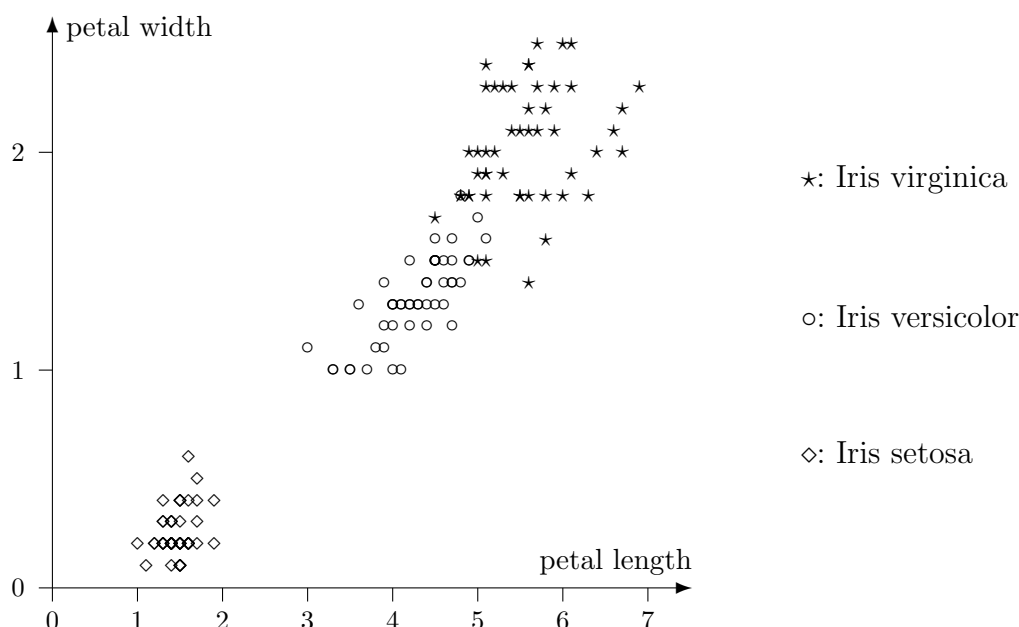
5. Übungsblatt

Aufgabe 25 Entscheidungsbaume: Visualisierung

Betrachten Sie den folgenden Entscheidungsbaum für die Iris-Daten (siehe den Datensatz auf der WWW-Seite zur Vorlesung sowie das dort zur Verfügung gestellte Entscheidungsbaum-Lernprogramm). Die Iris-Daten beschreiben Iris-Blumen durch ihre Blütenblattlänge und -breite (petal length and width, in cm), ihre Kelchblattlänge und -breite (sepal length and width, in cm) und ihre Zugehörigkeit zu einer der drei Arten Iris setosa, Iris virginica und Iris versicolor.

```
dtree(iris_type) =
{ (petal_length|2.45)
  <:{ Iris-setosa: 50 },
  >:{ (petal_width|1.75)
    <:{ (petal_length|4.95)
      <:{ Iris-versicolor: 47, Iris-virginica: 1 },
      >:{ (petal_width|1.55)
        <:{ Iris-virginica: 3 },
        >:{ Iris-versicolor: 2, Iris-virginica: 1 }},
      >:{ Iris-versicolor: 1, Iris-virginica: 45 }}}
```

Visualisieren Sie diesen Entscheidungsbaum durch Einzeichnen der Regionen, in denen die einzelnen Klassen vorhergesagt werden, in das folgende Diagramm:



Vergleichen Sie das Ergebnis mit der Klassifikation durch einen naiven oder vollen Bayes-Klassifikator (siehe Folien zur Vorlesung)!

Aufgabe 26 Induktion von Entscheidungsbäumen

Bestimmen Sie aus dem folgenden Datensatz nach dem Beispiel der Vorlesung (siehe Folien) einen Entscheidungsbaum für das Attribut „Class“, das angibt, ob unter den gegebenen Witterungsbedingungen Golf gespielt wurde! (Wählen Sie bei gleicher Güte zweier Attribute das in der Tabelle weiter links stehende.)

Outlook	Temp. ($^{\circ}F$)	Humidity (%)	Windy?	Class
sunny	85	85	false	Don't Play
sunny	80	90	true	Don't Play
overcast	83	78	false	Play
rain	70	96	false	Play
rain	68	80	false	Play
rain	65	70	true	Don't Play
overcast	64	65	true	Play
sunny	72	95	false	Don't Play
sunny	69	70	false	Play
rain	75	80	false	Play
sunny	75	70	true	Play
overcast	72	90	true	Play
overcast	81	75	false	Play
rain	71	80	true	Don't Play

Aufgabe 27 Induktion von Entscheidungsbäumen

Betrachten Sie den nebenstehenden Datensatz mit den drei booleschen beschreibenden Attributen A_1, A_2, A_3 und dem Klassenattribut C („w“ steht für „wahr“, „f“ für „falsch“). Wenn Sie mit dem in der Vorlesung beispielhaft vorgestellten Verfahren zur Induktion eines Entscheidungsbaum diesen Datensatz bearbeiten, wird der Entscheidungsbaum komplexer als nötig. (Wie ginge es einfacher?) Welches Problem der gierigen Attributauswahl zeigt dieses Beispiel?

A_1	f	w	w	w	f	f	f	w
A_2	f	w	f	f	w	f	w	w
A_3	f	f	f	w	w	w	w	f
C	f	f	w	w	w	f	w	f

Aufgabe 28 Entscheidungsbäume: Attributauswahlmaße

Berechnen Sie für die folgenden beiden Häufigkeitstabellen über den beschreibenden Attributen A, B und dem Klassenattribut C jeweils Informationsgewinn und χ^2 -Maß!

		A		
		a_1	a_2	a_3
C	c_1	9	4	3
	c_2	3	9	4
	c_3	4	3	9

		B		
		b_1	b_2	b_3
C	c_1	9	4	3
	c_2	6	6	4
	c_3	1	6	9

Wie könnte man das Auswahlverhalten der beiden Maße anschaulich beschreiben? (Hinweis: Beachten Sie die erste Zeile und letzte Spalte der beiden Tabellen.)