Maschinelles Lernen: Symbolische Ansätze

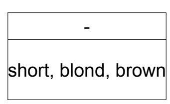
Übungsblatt für den 12.12.2006

Aufgabe 1

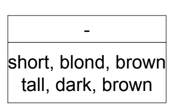
Vollziehen Sie die Berechnungen von ID5R für das Beispiel aus der Vorlesung nach.

height	hair	eyes	class
short	blond	brown	–
tall	dark	brown	–
tall	blond	blue	+
tall	dark	blue	–
short	dark	blue	–
tall	red	blue	+
tall	blond	brown	–
short	blond	blue	+

Lösung: Wir beginnen mit einem leeren Baum. Dieser wird nach dem Eintreffen des **ersten Beispieles** (short, blond, brown, –) durch ein einzelnes Blatt mit der Klassifikation – ersetzt, in dem das erste Beispiel gespeichert wird.

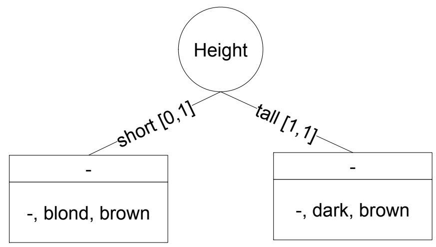


Das **zweite Beispiel** ändert nichts an der Struktur des Baumes, da das Beispiel durch das Blatt korrekt klassifiziert. Nur die Einträge innerhalb des Blattes werden geändert, d.h. man speichert jetzt auch das zweite Beispiel in ihm.



Da das **dritte Beispiel** positiv, reicht das Blatt nicht mehr aus und ein Test muß eingeführt werden. Die Implementierung des Verfahrens sieht vor, daß in Blättern keine

Häufigkeiten (der Attributwerte) gezählt werden, deshalb müssen wir zunächst eine beliebigen Test auswählen. Wir entscheiden uns für den Test *Height*. Damit erhalten wir zunächst den folgenden Baum.



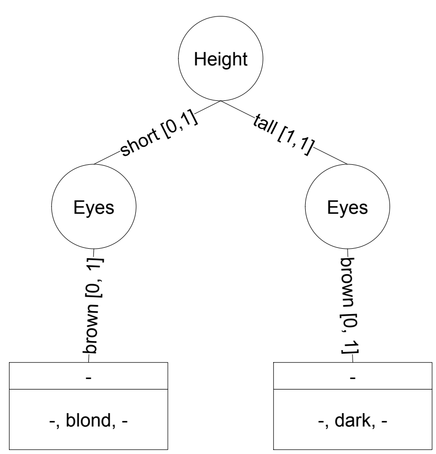
In dem eben eingeführten, inneren Knoten sind die absoluten Häufigkeiten bekannt, mit denen wir nun den besten Test auswählen können. *ID5R* verwendet hierfür als Bewertungsmaß *E-Score*, eine Variante des InformationGains, die auf die Berechnung der Entropie der Gesamtmenge (*Entropy(S)*) verzichtet, alle anderen Entropien aufaddiert und bei der der kleinste Wert ausgewählt wird (**Anmerkung**: *maximaler Information-Gain* ist äquivalent zu *minimalem E-Score*):

$$\textit{E-Score}(S, A) = \sum_{v \in Dom(A)} \frac{S_v}{S} Entropie(S_v)$$

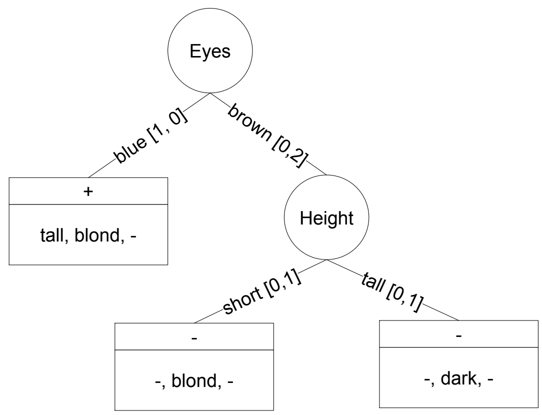
Wir erhalten für die drei möglichen Tests die folgenden *E-Scores*:

- $\textit{E-Score}(Eyes) = 0$
- $\textit{E-Score}(Hair) = 0,667$
- $\textit{E-Score}(Height) = 0,667$

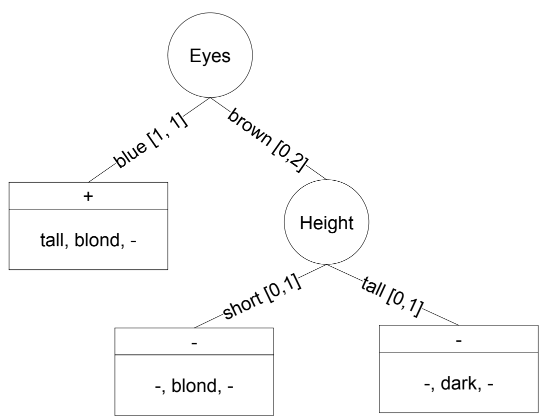
Demnach wird *Eyes* als bester Test ausgewählt. Damit wir *Height* durch *Eyes* ersetzen können, führen wir *Eyes* in den Baum. Damit erhalten wir den folgenden Baum:



Jetzt können wir die beiden Tests tauschen. Der Zweig, auf dem das dritte Beispiel weitergeleitet wird, benötigt keinen weiteren Test. Demnach sind der Baum nach dem Eintreffen des dritten Beispiels wie folgt aus:



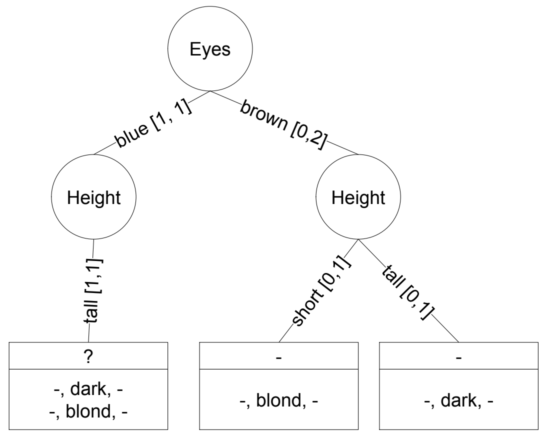
Nach Eintreffen des **vierten Beispiels** ändern sich die absoluten Häufigkeiten im Wurzelknoten *Eyes*.



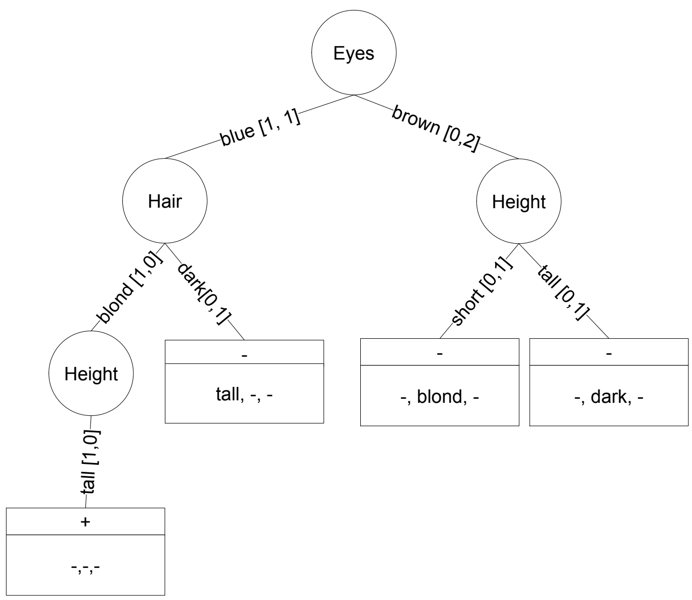
Wir berechnen nun, welcher Test jetzt am besten ist.

- *E-Score*$(Eyes) = 0,5$
- *E-Score*$(Hair) = 0,5$
- *E-Score*$(Height) = 0,689$

Eyes ist noch immer der beste Test in der Wurzel. Wir reichen das Beispiel eine Stufe weiter. Da es sich hier um ein Blatt mit positiver Klassifikation ist, aber das Beispiel negativ ist, müssen wir einen beliebigen Test (*Height*) einführen:



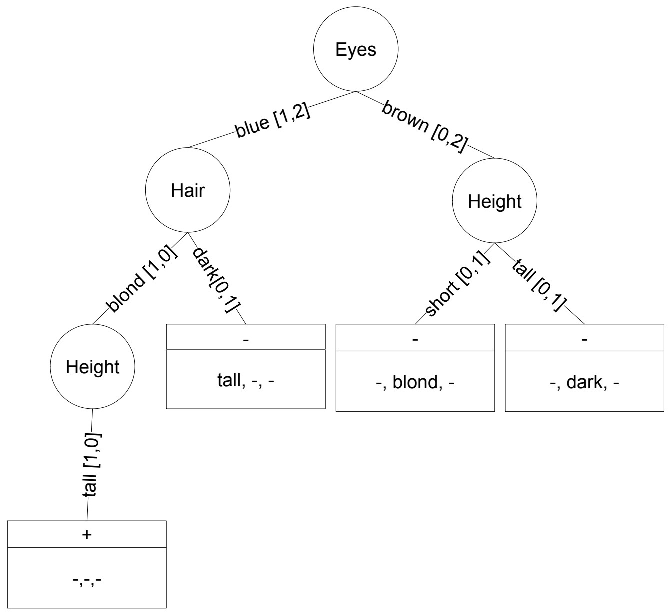
Wie man sieht, trennt der Test *Height* nicht die Beispiele. Für den Test *Hair* ist dies jedoch der Fall, deshalb führen wir diesen Test ein und vertauschen ihn mit *Height*:



Anmerkung: die zugehörige Berechnung sieht wie folgt aus:

- *E-Score*$(Hair) = 0$
- *E-Score*$(Height) = 1$

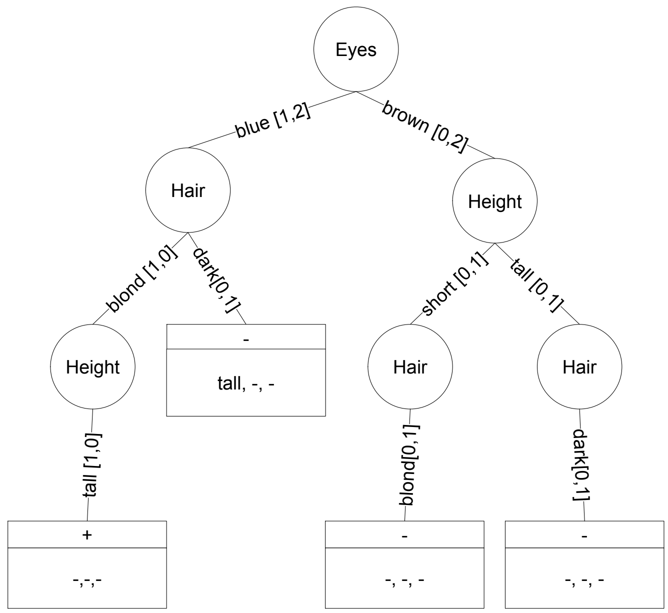
Das **fünfte Beispiel** ändert wiederum die absoluten Häufigkeiten in der Wurzel:



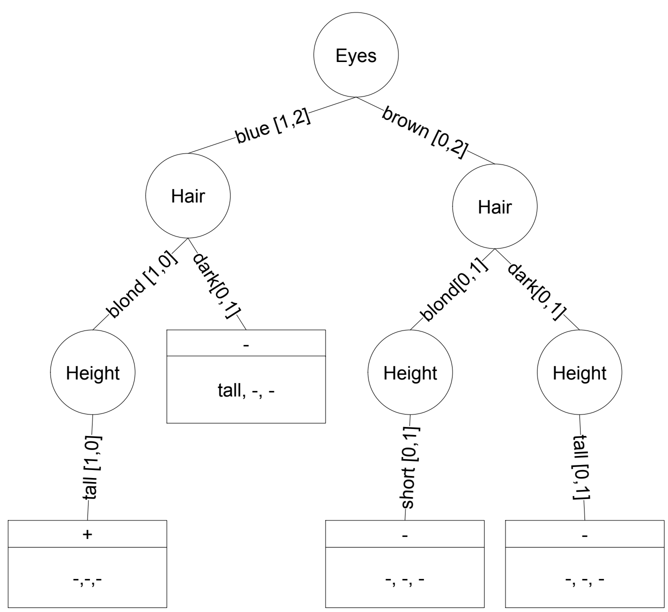
Mit diesen Werte berechnen wir erneut den besten Test in der Wurzel:

- *E-Score*$(Eyes) = 0,551$
- *E-Score*$(Hair) = 0,4$
- *E-Score*$(Height) = 0,551$

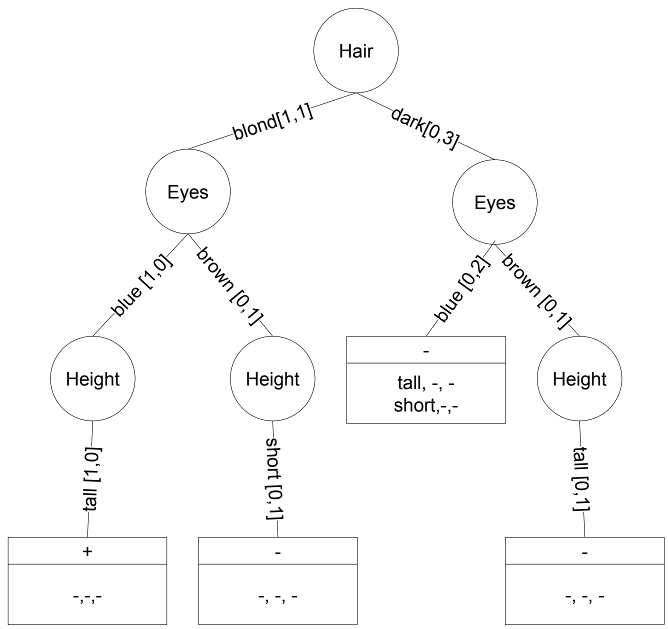
Hair ist nun der beste Test. Die Wurzel des linken Teilbaumes ist bereits *Hair*, deshalb muß dieser Teilbaum nicht verändert werden. Betrachten wir den rechten Teilbaum, in den der Test *Hair* an den Blättern eingeführt werden muß:



Jetzt können wir *Hair* mit *Height* vertauschen, dadurch erhalten wir den folgenden Baum:



Anschließend können *Height* mit *Eyes* vertauschen. Der Baum sieht nach Eintreffen des fünften Beispiels dann wie folgt aus:

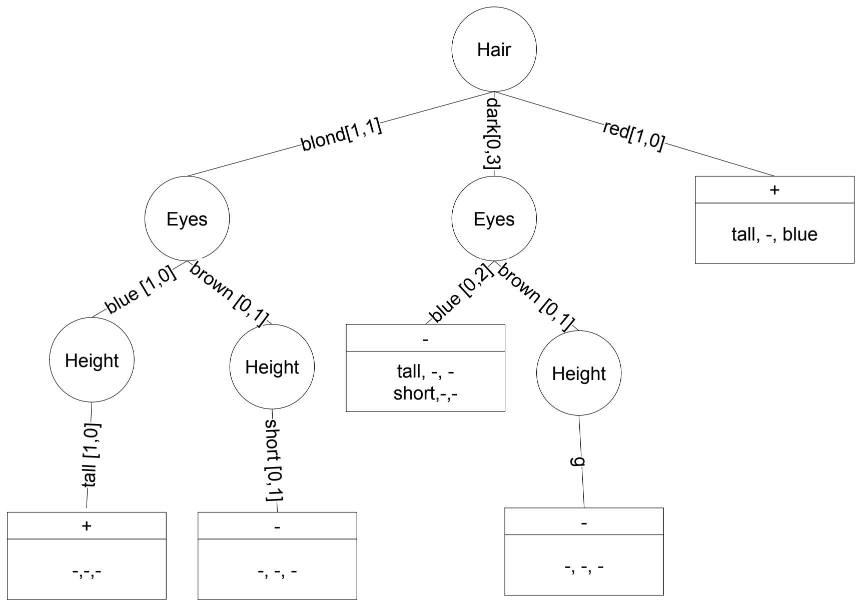


Auf dem Weg des Beispiels durch den Baum ändern sich die bereits bestehenden Tests nicht.

Das **sechste Beispiel** ändert erneut die absoluten Häufigkeiten in der Wurzel (diesmal verzichten wir auf eine Skizze). Mit den resultierenden Werten berechnen wir den besten Test in der Wurzel des Baumes.

- *E-Score*$(Eyes) = 0,667$
- *E-Score*$(Hair) = 0,333$
- *E-Score*$(Height) = 0,667$

Hair ist noch immer der beste Test. Für das Beispiel gilt *Hair* = *red*, aus diesem Grund müssen wir an der Wurzel einen neuen Zweig wachsen lassen. An diesen wir dann ein Blatt mit positiver Vorhersage angehängt:



Für die **letzten beiden Beispiele** ändert sich der Baum nur geringfügig. Die Änderungen können den Vorlesungsfolien entnommen werden.

Aufgabe 2

Eine Möglichkeit, aus unvollständigen Beispielen zu lernen, ist, Wahrscheinlichkeiten für jeden möglichen Attributwert zuzuweisen und Anteile der Beispiele in die Teilbäume zu propagieren (siehe *Entscheidungsbäume* Folie 34).

Diskutieren Sie diese Idee für ID5R.

- Welche grundlegenden Probleme tauchen hier auf?

  Lösung:

  1. Da in dem Szenario die Beispiele nacheinander verarbeitet werden, sind die Anteilsabschätzungen ungenau. Da man die Anteile nur anhand der vorherigen Beispiele ermitteln kann, hat man im Extremfall (das erste Beispiel ist bereits unvollständig) keine Möglichkeit überhaupt Anteile zu errechnen.
  2. Ein weiteres Problem ergibt sich daraus, dass mit jedem Eintreffen eines neuen Beispiels die Anteile der Attributwerthäufigkeiten aller unvollständigen Beispiele geändert werden müssen. Um allerdings herauszufinden welche Beispiele bisher überhaupt eingetroffen sind, muss man in jedem einzelnen Blatt nachschauen welche Beispiele dort gespeichert sind und wel-

chen Wert sie für das nicht vorhandene Attribut des unvollständigen Beispiels haben. Dazu muss man den kompletten Baum absuchen.

3. Da sich nun beim Eintreffen eines neuen Beispiels eben alle Attributwerthäufigkeiten von allen unvollständigen Beispielen geändert haben, muss in jedem einzelnen Knoten nachgesehen werde, ob hier noch der beste Test gewählt ist. Daher müssen sämtliche Pfade des Baumes durchwandert werden und nicht mehr nur der des aktuellen Beispiels. Außerdem kommt es viel häufiger zu Rebalancierung des Baumes.
4. Wenn ein neues unvollständiges Beispiel eintrifft, muss man nicht mehr nur einen Pfad aktualisieren, sondern alle Pfade des unbekannten Attributs. Ein unvollständiges Beispiel ist natürlich in mehreren Blättern (anteilig) gespeichert.

- Wie kann man mit diesen Problemen umgehen?

1. Für das erste Problem gibt es so gesehen keine Lösung, da man während des Lernens zu keinem Zeitpunkt Zugriff auf alle Daten haben wird (eine Annahme des inkrementellen Lernens) und daher die Attributwerthäufigkeiten immer nur anhand der bereits gesehenen Beispiele errechnen kann.
2. Für die Lösung dieses Problems könnte man eine Tabelle mit den relativen Attributwerthäufigkeiten irgendwo im Baum mitführen, so dass man in dieser beim Eintreffen eines neuen Beispiels direkt die neuen Häufigkeiten berechnen könnte.
3. In diesem Fall könnte man beispielsweise nur alle n Beispiele eine Umstrukturierung des Baumes aufgrund von den Änderungen der Attributwerthäufigkeiten der unvollständigen Beispiele durchführen. Ansonsten muss man den höheren Berechnungsaufwand in Kauf nehmen.
4. Da die Anteile der Attributwerte in die Bäume propagiert werden, kann man hier keine Abhilfe schaffen.

Eine Zusammenfassung dieser Aufgabe ist zusätzlich in der Musterlösung zur Klausur Maschinelles Lernen: Symbolische Ansätze aus dem Wintersemester 2005/06 zu finden (Aufgabe 3-c).

Aufgabe 3

Gegeben sei eine Beispielmenge analog dem Wetterbeispiel mit 1000 Datensätzen, die zeitlich geordnet sind (jeweils 1 Beispiel / Tag).

a) Wie würden Sie die Verfahren *Aging* und *Windowing* implementieren?

Lösung: Die hier vorgeschlagenen Lösungen sind bei beiden Aufgaben nur exemplarisch und nicht vollständig. Sie sollen nur einen Überblick geben.

Bei *Aging* muss man sich überlegen, wie stark man das Alter der Daten ins Gewicht nimmt. Verwendet man beispielsweise die Formel $1/_{Alter}$, so fallen Beispiele, die älter sind, nicht so stark ins Gewicht wie neuere Beispiele. Auch hier kann man eine Grenze einführen, ab der die Beispiele, deren Gewicht einen Schwellwert unterschreitet, aus der Liste gestrichen werden.

Bei *Windowing* ist die wichtigste Überlegung, wie groß man das Fenster wählen sollte. Hierfür eignen sich Methoden wie *Sliding Window*. Bei diesem wird ein Fenster immer ein Stück weit nach vorne verschoben, so dass an der unteren Grenze ältere Daten herausfallen, an der oberen Grenze neue Daten hinzukommen und in der Mitte ein Teil der alten Daten verbleibt. So kann man die Fenstergröße an die aktuelle Situation dynamisch anpassen. Üblicherweise würde hier das Fenster größer werden, je weniger Konzeptdrift vorliegt und kleiner je größer der Konzeptdrift ist.

b) Wie bekommen Sie heraus, ob Konzeptdrift vorliegt?

Unter Konzeptdrift versteht man die langsame Änderung eines Konzeptes, unter Konzeptshift hingegen die schnelle, abrupte Änderung. Als Beispiel könnte man sich als zu lernendes Konzept eine Funktion $f(x) = a \cdot x + b$ vorstellen, wo sich die Konstante b langsam erhöht. In diesem Fall würde Konzeptdrift vorliegen. Verändert sich die Funktion abrupt zum Beispiel zu $f(x) = a \cdot x^{10} + b$, so würde Konzeptshift vorliegen. Generell ist hier jedoch keine genaue Abgrenzung möglich.

Um herauszufinden ob Konzeptdrift vorliegt, kann man beispielsweise eine Änderung im Erwartungswert berechnen (Varianzen auf den Attributwerten) und dann über Signifikanztests feststellen, ob es zu Veränderungen gekommen ist, die groß genug sind, um Konzeptdrift vorhersagen zu können. Eine andere einfache Möglichkeit wäre, Konzeptdrift immer dann anzunehmen, wenn die Vorhersagegenauigkeit geringer wird.