

Decision Tree Pruning (Pen & Paper)

Figure 1 shows a Decision tree. Which branch can be pruned such that the prediction accuracy of the given test sample does not decrease (the prediction accuracy is defined as the ratio of correctly predicted samples divided by the number of all samples). Which is the minimal prediction tree such that the prediction accuracy of the test sample is still above “0.8”?

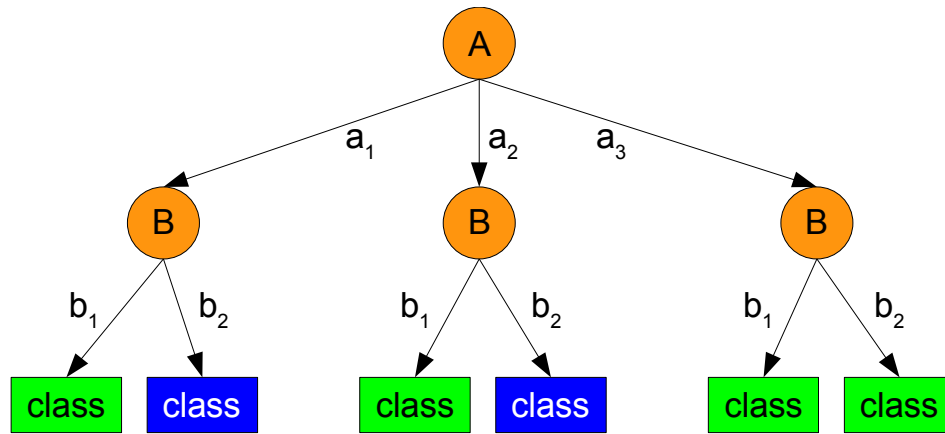


Fig. 1:

Use the following test samples to measure the prediction accuracy:

- {a1,b1,green}, {a1,b2,blue}, {a2,b1,green}, {a2,b2,blue}, {a3,b1,green}, {a3,b2,green}

Decision tree for prediction (Pen & Paper)

Create a minimal decision tree for the following training sample set:

A	B	C	D	E	Class
gut	rot	A-Team	1	Rose	green
gut	blau	Star Trek	3	Lilie	blue
schlecht	blau	Scrubs	1	Lilie	blue
mittel	rot	Batman	3	Tulpe	green
gut	schwarz	A-Team	1	Rose	green

Using your tree, predict the class for the following samples:

- {schlecht,blau,Scrubs,1,Rose}
- {mittel,rot,Batman,3,Lilie}
- {schlecht,schwarz,Star Trek,2,Veilchen}

You may use the entropy measure to decide which Attribute splits the samples, however, this is not a requirement.