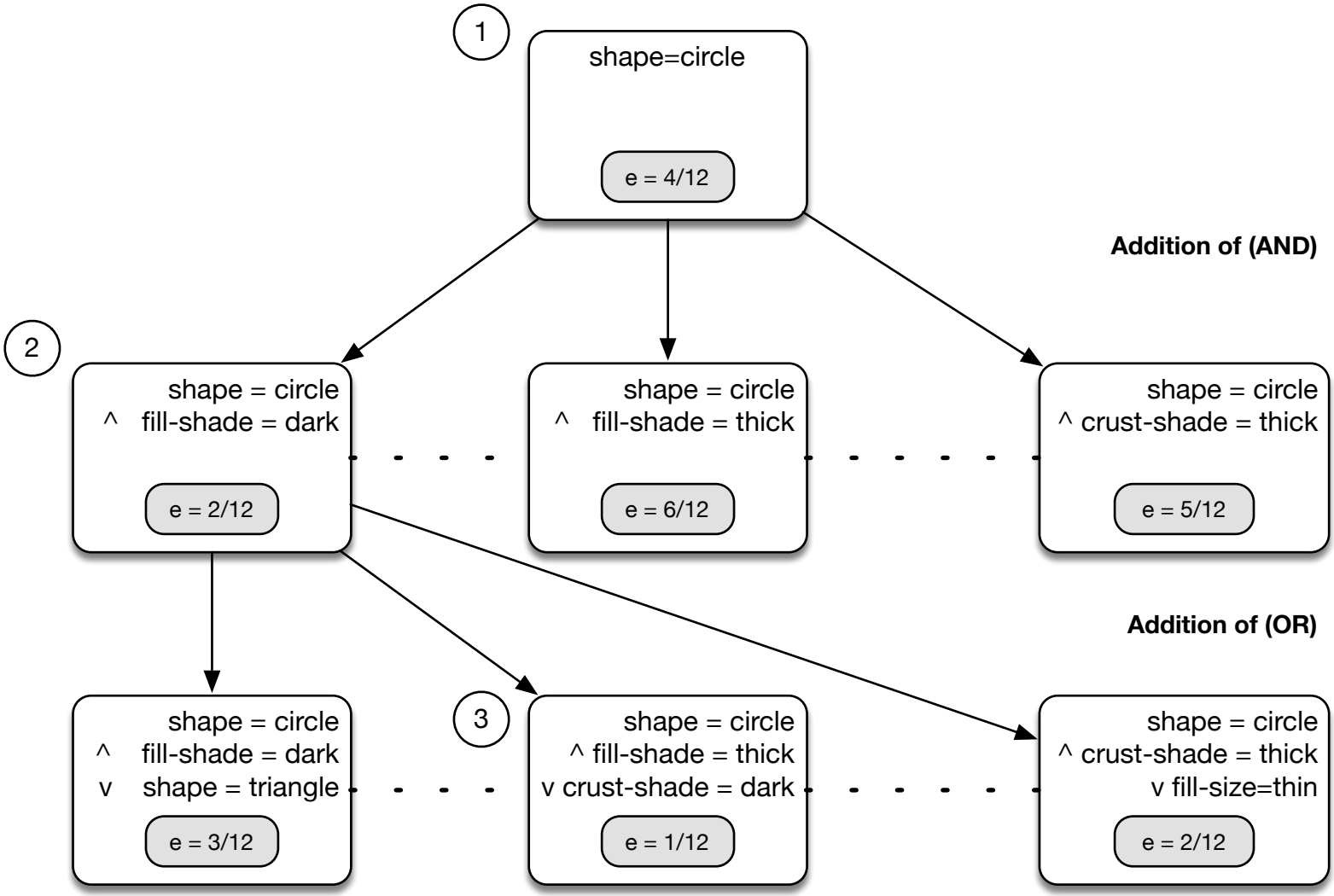
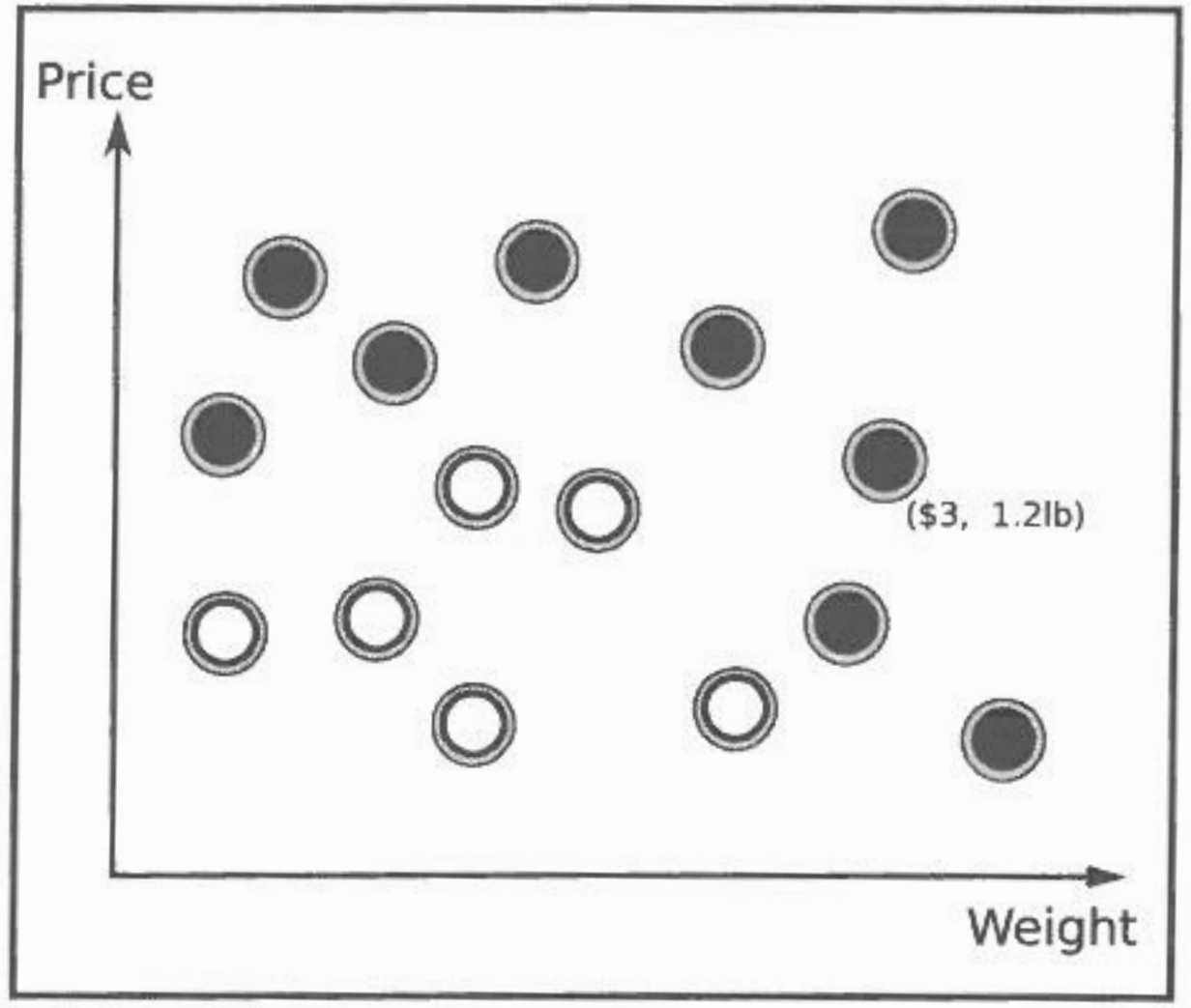
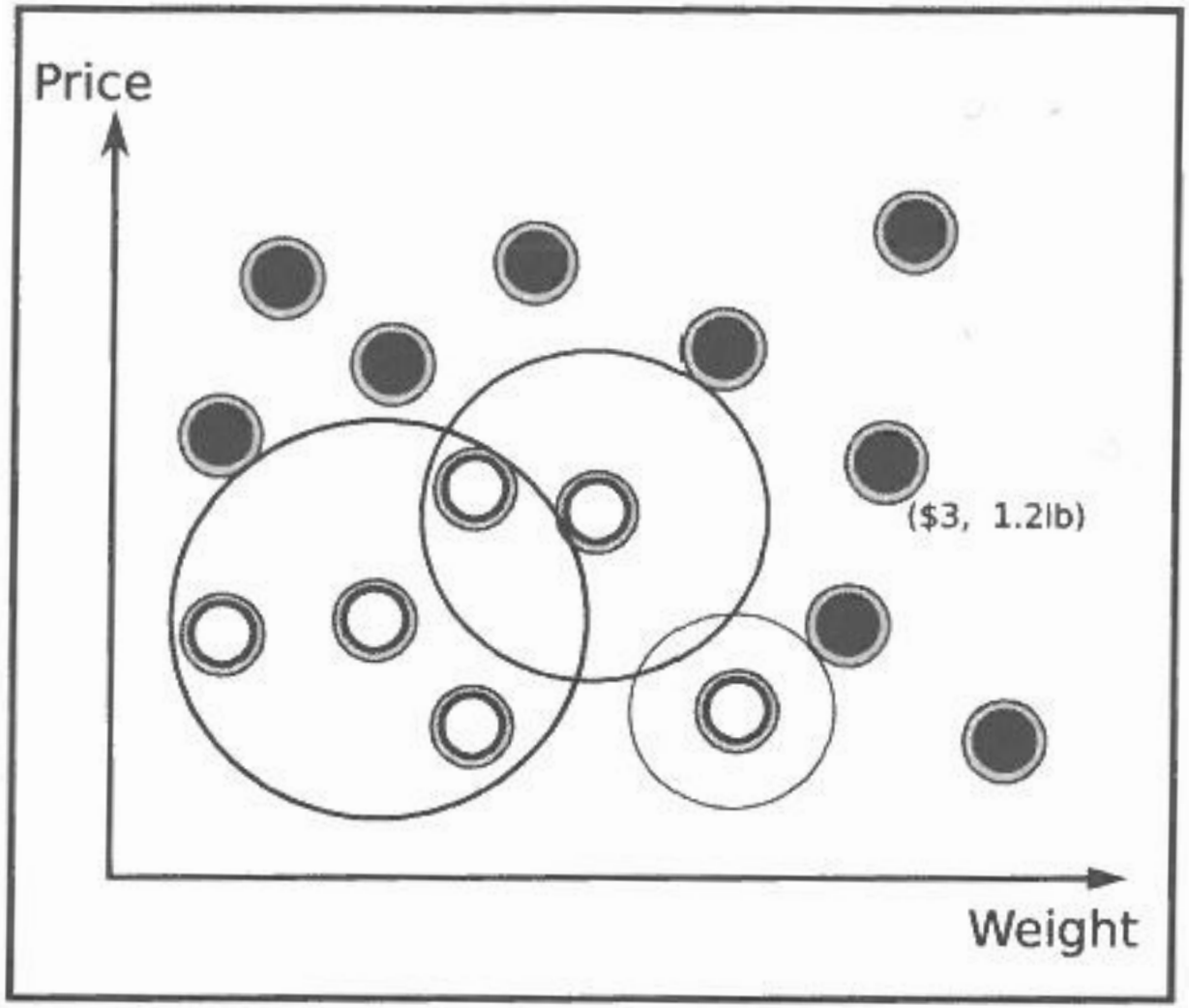
Hill-Climbing in the “pies” domain



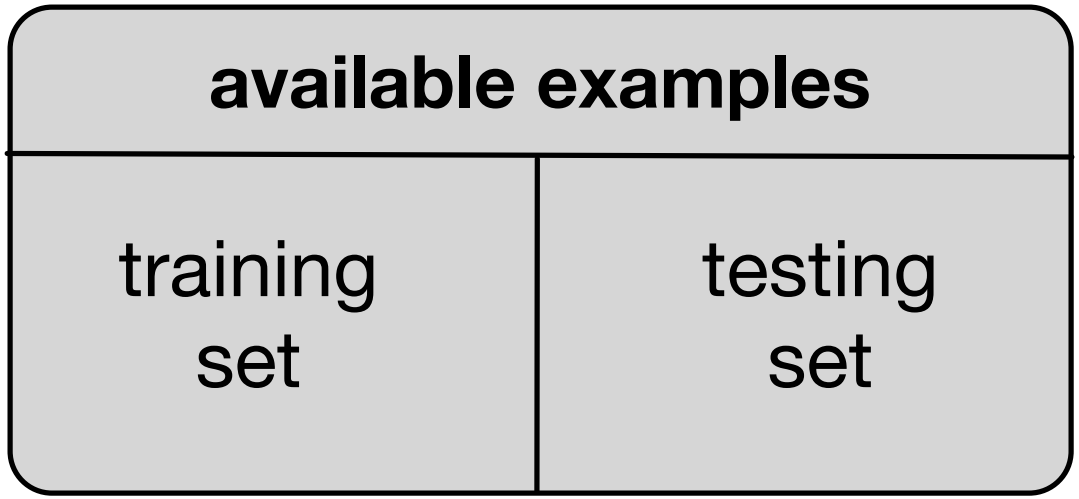
Domain with Continuous Attributes



Domain with Continuous Attributes: Circular Classifier



Classifier's Performance



e.g. select 8 from 12 examples as training data randomly. Then testing the classifier on the remaining four.

Discussion

- How can we estimate the error rate on examples that have not been seen during learning?
- Why is error rate usually higher on the testing set than on the training set?

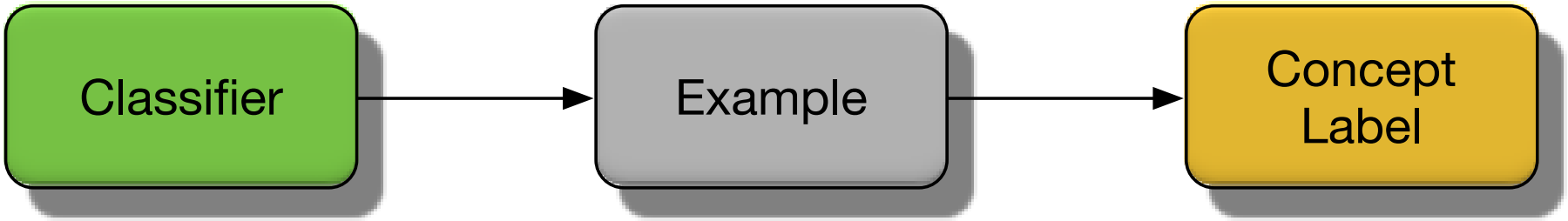
Difficulties with Available Data

- Irrelevant attributes
- Missing attributes
- Redundant attributes
- Attribute- value noise
- Class-label noise

Training and Application

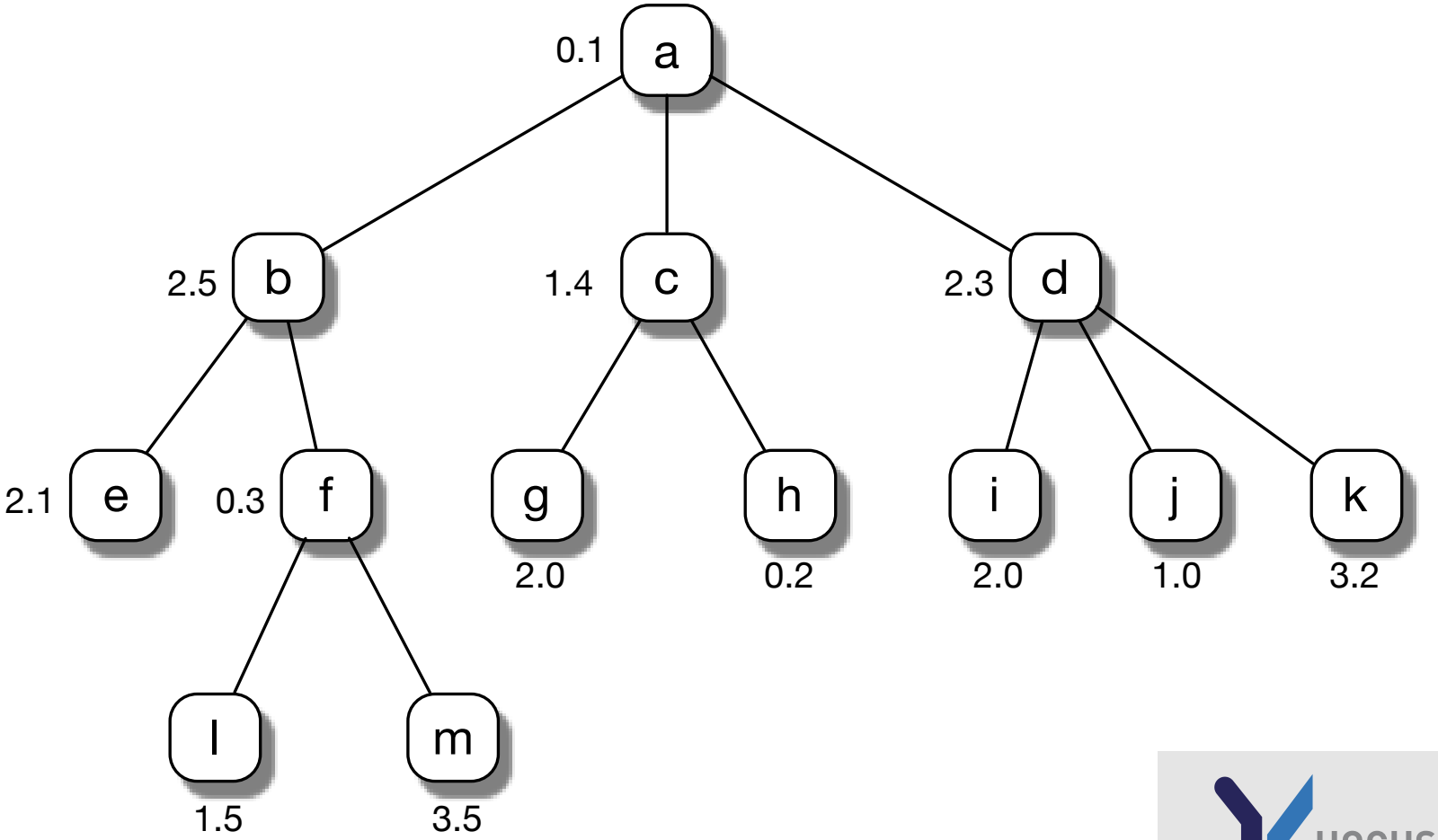
Learning:

Application:



Exercise

Determine the order in which these search states are visited by hill-climbing algorithm?



Exercise

Discuss in a group of two.

Take a class that you think is difficult to describe — for instance, the recognition of a complex biological object (oak tree, ostrich, etc.) or the recognition of music genre (rock, folk, jazz, etc.). Suggest the list of attributes to describe the training examples. Are the values of these attributes easy to obtain? Which of the problems do you expect will complicate the learning process?