Assignment 1

Bendik Nordeng

September 2019

1 Theory

1. Concept learning is about inferring a boolean-valued function from training examples of its input and output. One example could be to distinguish birds from all other animals.

2. Function approximation is about choosing a function among a well-defined class that closely matches a target function. Function approximation is needed because learning algorithms sometimes may be too complex for learning. Hence, we can only expect to acquire only some approximation to the target function.

3. The inductive bias of a learning algorithm is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered. For a decision tree, the inductive bias will be the decision boundaries. As for the candidate elimination algorithm, the inductive bias is that the target concept is contained in the given hypothesis space.

4. Overfitting occurs when a model is too specific in it's learning of a concept. For a decision tree this could be an effect of too much depth. A model that overfits will perform well on the training set (often 100% accurate) and poorly on the test set. Underfitting is when a model is too general, maybe learning too few attributes. A model that underfits will perform poorly on both training- and test set A well performing model will find a good balance between over- and underfitting. A *validation set* provides an unbiased evaluation of a model fit on the training set. This can be used to tune hyperparamaters such as decision tree depth or hidden layers in a neural network. Cross validation is the process of using different parts of the dataset for training and testing. This mitigates overfitting because the model can learn from all of the data.

5.
   - $S_0 = \{< \emptyset, \emptyset, \emptyset, \emptyset >\}$
     $G_0 = \{<?, ?, ?, ? >\}$
   - $S_1 = \{< Female, Back, Medium, Medium >,$
     $< Female, Neck, Medium, High >\}$
     $G_1 = \{<?, ?, ?, ? >\}$

- $S_2 = \{< Female, ?, Medium, ? >\}$
  $G_2 = \{<?, ?, ?, ? >\}$
- $S_3 = \{< Female, ?, Medium, ? >\}$
  $G_3 = \{<?, ?, Medium, ? >\}$
- $S_4 = \{<?, ?, ?, ? >\}$