Dan Scarafoni
homework 19

1.

a. Unsupervised learning- the agent learns patterns in input with no feedback.
b. Supervised learning- the agent receives some example input-output pairs which it uses to help derive the function
c. Reinforcement learning- the agents receives rewards or punishments in response to its actions.

2.

a. a is the simplest hypothesis, as it's hypothesis space H only has two dimensions- real numbers m and b. C is the next simplest, as it's H is three dimensional (a,b,c as real numbers). B is the most complex as its H has four dimensions- a,b,c,d.

b. A fits the equation to the least degree. This is evident as none of the data points even graze the fit line. B and C fit comparably, as all data points fall on the fit lines. For this reason, both B and C are consistent hypotheses.

c. A is very generalizable. Because a linear equation only requires two points in order to solve for an equation, the third point provides additional data to help further validate the hypothesis. B, on the other hand, is not very generalizable at all. It has four unknown variables and only three points. It does not even have enough data points to find a regression of that magnitude. C is more generalizable. Only having three unknown variables, a fit equation can be soundly derived.

3. There is a trade-off between complex hypotheses that fit well and simpler ones that are more generalizable. More complex hypotheses do not generalize well. Further, there are computational limitations to finding good hypotheses. Sufficiently complex hypothesis spaces are nearly (if not completely) impossible to calculate.

4.

$$(Patrons = some) \lor ((Patrons = full) \land (Hungry = yes) \land ((Type = french \lor burger \lor (thai \land Fri/Sat = yes)))) \rightarrow Eat$$

5. It means that the examples remaining have the same description, but not the same classification. This can be caused by an error in the data (noise), a non deterministic domain, or by an unobserved attribute that would explain the difference.

6. A learning curve, which can be produced by generating different sized test and training sets from examples and testing their accuracy vs the size of the training set, can help evaluate the accuracy of an algorithm. By calculating the information gain of choosing a certain attribute for a decision learning tree, a more efficient (shallower) tree can be produced. Further, k fold cross-validation can be used. In this method, data is divided into k parts, each of roughly equal size. Each section is used as the test set exactly once, and the rest of the groups are used as the training set. The average error across all trials is computed.