**CSE 5693 Machine Learning**

**HW4 Genetic Algorithm**

Josias Moukpe

Written Assignment

1. **9.4**

**Consider applying GAS to the task of finding an appropriate set of weights for an artificial neural network (in particular, a feedforward network identical to those trained by BACKPROPAGATION. Consider a 3 x 2 x 1 layered, feedforward network. Describe an encoding of network weights as a bit string, and describe an appropriate set of crossover operators. Hint: Do not allow all possible crossover operations on bit strings. State one advantage and one disadvantage of using GAs in contrast to BACKPROPAGATION to train network weights**

Answer:

A 3x2x1 network means there’s 6 + 2 + 2 + 1 = 11 weights to learn by the GA. One possible encoding of the weight could be a floating-point number binary encoding of the weights concatenated into a string. That floating-point encoding could be IEEE 754 standard or simply decimal to binary conversion with support for tenths, hundredths, and possibly thousands (.000). The IEEE standard has the advantage of representing more values while the simple conversion has a smaller search space.

Not all crossover operations will be accepted since we want to maintain the size of the bit string so all the weights are always accounted for. All the crossover operations that maintain bit string size are acceptable. One example could be a uniform crossover operator with a random mask.

One advantage of using GA is that there are fewer hyperparameters to work with, making it simpler to operate.

One disadvantage is the encoding of floating points, backpropagation is much better equipped to search infinite continuous spaces than GAs since backpropagation is guided search in the descent direction. The GA will take longer to converge and will be more likely to converge to a local minimum.

1. **10.1**

**Consider a sequential covering algorithm such as CN2 and a simultaneous covering algorithm such as ID3. Both algorithms are to be used to learn a target concept defined over instances represented by conjunctions of n boolean attributes. If ID3 learns a balanced decision tree of depth d, it will contain 2d - 1 distinct decision nodes, and therefore will have made 2d - 1 distinct choices while constructing its output hypothesis. How many rules will be formed if this tree is re-expressed as t a disjunctive set of rules? How many preconditions will each rule possess? How many distinct choices would a sequential covering algorithm have to make to learn this same set of rules? Which system do you suspect would be more prone to overfitting if both were given the same training data?**

Answer:

1. **10.3**

**Refine the LEARN-ONE-RULE algorithm of Table 10.2 so that it can learn rules whose preconditions include constraints such as nationality belong to set {Canadian, Brazilian}, where a discrete-valued attribute is allowed to take on any value in some specified set. Your modified program should explore the hypothesis space containing all such subsets. Specify your new algorithm as a set of editing changes to the algorithm of Table 10.2.**

1. **10.6**

**Apply inverse resolution to the clauses C = R(B, x) v P(x, A) and C1 = S(B, y) v R(z, x). Give at least four possible results for C2. Here A and B are constants, x and y are variables.**

Answer:

L = S(B, y)

C2 = - S(B, y) v P(x, A) with z/B

Another example for C2 is

C2 = - S(k, y) v P(x, l) where k and l are additional variables

1. **From testIrisSelection in the programming assignment, compare the three selection strategies. Plot test set accuracy against number of generations and discuss your observations.**
2. **From testIrisReplacement in the programming assignment, plot test set accuracy against replcaement rate (r) and discuss/explain your observations.**