CSE473 Homework #4
Due at 10:30 AM December 11th in Class on Paper
Name:

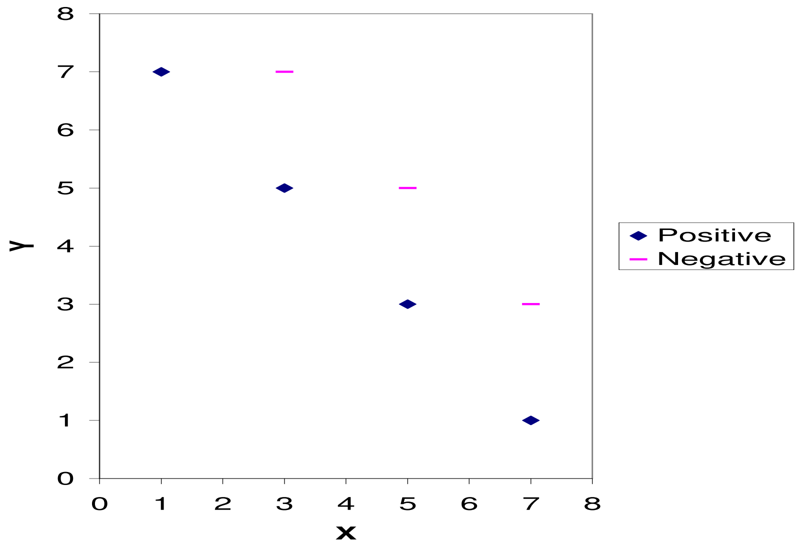
Model ensembles

A significant disadvantage of model ensembles is that an ensemble is significantly more complex and harder to understand than a single model. Propose a computationally efficient method for reducing a set of bagged decision trees to a single tree that is approximately equivalent to it.

Instance-based learning

Voronoi Diagram

Draw the Voronoi diagram corresponding to the following training set:



The instance space is the square

$$0 \leq X \leq 8, 0 \leq Y \leq 8$$

Positive examples are located at

- (1,7)
- (3,5)
- (5,3)
- (7,1)

Negative examples are located at

- (3,7)
- (5,5)
- (7.3)

Error Rate

If the true frontier between positive and negative examples is the line

$$Y = 9 - X$$

with the positive class on the left, and the examples are uniformly distributed in the square above, what is the expected error rate of the nearest-neighbor classifier with Euclidean distance trained on the examples above?

Genetic algorithms

Consider a genetic algorithm that uses fitness-proportionate selection and single-point crossover. Half the population is replaced by crossover at each step, and there is no mutation. Suppose the current population is composed of four hypotheses h_1, h_2, h_3 and h_4, with the fitness of the ith hypothesis being i (e.g., $fitness(h_3) = 3$). If the first bit is 0 in the even hypotheses and 1 in the odd ones, what is the expected number of hypotheses with 1 in the first bit in the next generation?

Neural networks

DNF

Write in pseudo-code a function that inputs a logical formula in DNF (disjunctive normal form) and outputs a neural network that implements it, with each neuron being represented as a list of inputs and their respective weights. The firing threshold for each neuron should be greater than zero (as opposed to greater than or equal to zero).

CNF

How would you modify this function to input instead a logical formula in CNF (conjunctive normal form)?