LAST (FAMILY) NAME: _____FIRST NAME: _____
12 questions, 5 points each, total 60.

Question 1. You have a dataset with 1000 data points which you want to use to train a kNN classifier. You setup the experiment as follows: you train 5 kNN classifiers (k={1,3,5,10,25}) using all the data points. Then you randomly select 300 data points from the 1000, and classify them using each of the 5 classifiers. Which classifier will come out as the best one and what do we call this phenomenon?

Question 2. The majority classifier (classify a test point by the majority class of the training set, regardless of the features of the test point) is a special case of a kNN classifier. Describe the relationship between them.

Question 3. In clustering, one way of choosing the number of clusters k is to use regularization. A common way of doing this is to use the Schwarz criterion which picks k by minimizing: distortion + λ * D * k * log N where D is the dimension of the problem, k is the number of clusters and N is the number of data points. λ is parameter which you have to specify. What is the influence of λ on the optimal k, when λ approaches 0 and infinity?

Question 4. Consider the following one-dimensional dataset: {1,2,4,8,16,32} and let us assume we measure distance between points using the standard Euclidean distance. Does it produce different clusters whether we use single linkage or complete linkage when we apply hierarchical clustering?

Question 5. I have a book with 1024 pages. I randomly flip to a page and ask you to determine the page number. You can only ask questions with yes/no answers.

a) How many questions do you have to ask to guarantee an answer, if you design your questions optimally?
b) How do you design your questions?
c) If you observe that I tend to flip to the first half of the book, does it make the number of questions needed: smaller, the same, or bigger in theory?

Question 6. You are developing a route planning algorithm. Since the algorithm has to run on a car computer, you only have very limited amount of memory available. Which algorithm would you prefer: depth first or breadth first search? Briefly argue why.

Question 7. If h1 and h2 are admissible search heuristics, which are admissible heuristics below?
A. 0.1*h1 + 0.9*h2 B. min(h1,h2) C. min(h1,2*h2) D. min(h1,0.5*h2) E. h1– h2

Question 8. Show the A* search procedure on the following graph, starting from state A. G is the goal state. All heuristic values are zero, and all edges cost one:
A ---> B <---- G

Question 9. What is the relationship between alpha-beta pruning and minimax search?

Question 10. Given the following dataset:

x	y
1	+
0	-

What is the entropy H(y)? What is the conditional entropy H(y|x)? Use bits as the unit.

Question 11. We use gradient descent to find the minimum of the function f(x) = |x| with step size α=1. If we start from the point x0=-5 (minus 5), what is the point we arrive at in one step?

Question 12. Compute the minimax value of the following game.

	Min-I	Min-II
Max-I	30	20
Max-II	10	10

[Start your answers here]