

1. The kNN classifier with $k=1$ will come out as the best one since we are choosing our test points from the training points, the 1NN classifier will have remembered the correct label and the 1NN classifier will achieve 0 error rate. We call this overfitting.
2. If we have a dataset with n datapoints, then the n -nearest neighbor classifier will always use all datapoints in the dataset to classify new points. Since nearest neighbor classifiers use a majority voting scheme, the n NN classifier is the majority classifier.
3. When λ is 0, we are not penalizing having a large number of cluster centers. There is a trivial clustering which achieves zero distortion by putting a cluster center at every datapoint. When λ goes to infinity, the penalty for one extra cluster is going to dominate the distortion and we will have to do with the least amount of clusters possible which is $k=1$.
- 4.

5. Since 1024 is 2^{10} , which we can represent as a 10 character string over the alphabet 0,1. For question i , one can ask whether the bit is 1 or not. Thus asking 10 questions is enough to guarantee an answer. If we know the page is in the first half of the book, this means the page number will be at most 512 which in our representation (if we let page 1 correspond to 0000000000) means that the page number will start with a 0. We thus only have to ask 9 questions.
6. Since breadth first search requires a very large amount of memory (it increases exponentially at each level when we have a constant branching factor) we would prefer depth first search.
7. All are since the only requirement is that a heuristic must underestimate the actual cost to the goal.
8. A ($h=0, g=0$)
B ($h=0, g=1$)
end, no goal found
9. Alpha-beta pruning is essentially the same procedure as mini-max except that it is a little smarter about which nodes to evaluate. Essentially alpha-beta pruning keeps a little extra bookkeeping to be able to prune away branches that cannot possibly affect the final decision. It is therefore a little more efficient in terms of time complexity.
10. $H(Y) = 1, H(Y|X) = 0$.
11. We end up at $x=-4$ since the gradient at $x=5$ is 1.
12. The minimax value will be 20 since when Max plays strategy I and Min plays strategy II, no player can improve upon its score without the other decreasing her score.