Machine Learning *Angjoo Kanazawa (Kim) 111964222*

HW01: Basic concepts and geometry

Hand in at: `http://www.cs.utah.edu/~hal/handin.pl?course=cs726`. Remember that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running "`pdflatex hw00.tex`".

1. Give two reasons why simply memorizing training data and doing table lookups is insufficient for learning.

   Because you may not be able to do a look up on all test data and because it won't generalize. The test data is likely to contain objects that weren't present in training data (with features that wasn't in the training), and therefore you won't be able to do a lookup. This method will only work for examples that we've already seen, and this isn't learning. We want to predict the future well, not recite.

2. In two sentences, explain the difference between a parameter and a hyperparameter (you may focus on the decision tree model).

   A hyperparameter let's you adjust your algorithms to behave in a certain way when we don't know *a priori* what the optimal value of that hyperparameter should be. A parameter is a value that we can compute/select to get optimal results from your algorithms.

3. For an overfit model, do you expect the training error $\hat{\epsilon}$ to be greater or less than the text error ϵ?

   I expect the training error to be less than the test error.

4. A learning algorithm should never look at test data. You shouldn't either. Why not (one sentence is fine)?

   I shouldn't either because if I see it I will know what to expect, and unconsciously modify the algorithm, turn the hyper-parameter what not and try to fit my algorithm to the test-data that I've seen.

5. What happens to a KNN classifier if the scale of one feature dominates (i.e., is 1000 times larger than) all the others? (One sentence is fine.)

   If one of the feature dominates, a K-NN classifier will ignore all of the other features. (The distance metric will not correctly tell the "real" distance between feature points, or look at features fairly).