

CS 4824 / ECE 4424 Homework 2 (Written Portion) Solutions

Question 1 (Comparing Classifiers) [6 points]

- Decision trees, since k-NN would have problems with constructing a proximity measure that can handle continuous as well as nominal attributes.
- Convert the nominal attribute (type of job) into 4 asymmetric binary variables (or 2 symmetric binary variables), each indicating the presence of a particular type of job.

Question 2 (Comparing Classifiers) [8 points]

- kNN – Since it looks only for local neighborhood, it will be able to learn the complex boundary between the classes. NB will have poor performance due to variable interaction. Decision trees will have difficulty as it can only find rectilinear splits.
- Naïve Bayes, as the attributes are not conditionally independent. K-NN and ANN can handle variable interactions

Question 3 (Comparing Classifiers) [5 points]

Decision Trees and Multi-layer ANN. Both are universal approximators, i.e., we can construct an arbitrarily complex model that can perfectly classify every training instance, provided that the class label for every combination of attribute values is unique.

Practice Questions (will not be graded)

Question 4 (Naïve Bayes)

- a) Yes b) No c) Yes d) yes

Question 5 (ANN)

- False. The gradients are computed starting from the output layer to the hidden layers, rather than the opposite direction.
- True. The activations at k th layer can be directly computed using the activations at $k+1$ th layer as inputs.
- True. If the training loss is 0, then the gradient at the output node is 0, and hence the gradients at all hidden nodes is 0.

Question 6 (ANN and Logistic Regression)

Similarities (choose any two):

1. Both use linear decision boundaries, and hence cannot classify non-linearly separable data.
2. Both minimize only the training loss and do not directly minimize model complexity.
3. Both can handle variable interaction (as long as the classes are linearly separable), even with high-dimensional attributes.
4. Both are unable to handle missing attributes.

Differences (choose any two):

1. Perceptron using sign activation function while logistic regression using sigmoid activation function.
2. Perceptron only outputs the class label for every instance but logistic regression additionally provides the posterior probability of the positive class.
3. Perceptron learning rule provides no theoretical guarantee of optimality and can end up in a local minima, while the optimization problem of logistic regression is convex and results in a global minima.

Question 7 (Naïve Bayes)

False. NB does not require independence among attributes, $P(A, B) = P(A) \times P(B)$, but only requires conditional independence among attributes given the value of Y .

Question 8 (Comparing Classifiers)

(a) Both decision tree and NB will do well on this data set because the distinguishing attributes have better discriminating power than noisy attributes in terms of entropy gain and conditional probability. k-NN will not do as well due to relatively large number of noise attributes, which will adversely effect the computation of similarity between two examples.

(b) k-NN will work the best due to the proximity of the examples of the same class to each other. NB does not work well for this data set since the attributes that determine the class boundaries are not independent. Decision tree will have to be large in order to capture the circular decision boundaries, and thus is not the ideal solution.