CptS 570 Machine Learning: Sample Questions

1. **Short Questions**

   Please keep your answers short (at most two sentences). For the True / False questions, please also provide a short justification.

   a) The ID3 decision tree learning algorithm is guaranteed to find the optimal decision tree (True / False)

   b) We would expect the support vectors to remain the same in general as we move from a linear kernel to higher order polynomial kernels (True / False)

   c) Why does the kernel trick allow us to solve SVMs with high dimensional feature spaces, without significantly increasing the running time?

   d) Your professor wants to automatically classify PhD applications into good/bad categories, and also wants to detect outliers (applications not in his research area) using density estimation. Will you recommend him a discriminative or generative classifier? Why?

   e) Your professor who is new to WSU wants to automatically classify PhD applications who are likely to become great researchers after they graduate. However, he has a small amount of training data. To create the most accurate classifier, will you recommend him a discriminative or generative classifier? Why?

2. **Decision Trees**

   Suppose we want to learn a decision tree for predicting whether a person is *happy* or *sad* based on the color of their shirt, whether they wear a tie, and the number of eyes they have.

   Table 1: Training data for decision tree learning.

Color	Tie	No. of Eyes	(Output) Emotion
Green	Yes	2	Sad
Green	No	2	Sad
Green	No	2	Sad
Blue	No	2	Sad
Blue	No	2	Happy
Red	No	2	Happy
Red	No	2	Happy
Red	No	2	Happy
Red	Yes	3	Happy

   a) What is the conditional entropy $H(Emotion|Tie = yes)$?

   b) What is the conditional entropy $H(Emotion|Eyes = 3)$?

   c) Which feature would be selected first (root of the tree) by the decision-tree learning algorithm?

   d) Draw the full decision tree (without any pruning) that would be learned from this data?

3. **Bayes Rule**

a) I give you the following fact: $P(A|B) = 2/3$. Do you have enough information to compute $P(B|A)$? If not, write "not enough information." If yes, compute the value of $P(B|A)$.

b) I give you the following facts: $P(A|B) = 2/3$, $P(A|\neg B) = 1/3$. Do you have enough information to compute $P(B|A)$? If not, write "not enough information." If yes, compute the value of $P(B|A)$.

c) I give you the following facts: $P(A|B) = 2/3$, $P(A|\neg B) = 1/3$, $P(B) = 1/3$. Do you have enough information to compute $P(B|A)$? If not, write "not enough information." If yes, compute the value of $P(B|A)$.

d) I give you the following facts: $P(A|B) = 2/3$, $P(A|\neg B) = 1/3$, $P(B) = 1/3$, $P(A) = 4/9$. Do you have enough information to compute $P(B|A)$? If not, write "not enough information." If yes, compute the value of $P(B|A)$.

4. **Parameter Estimation**

a) SVM classifier tend to be robust to irrelevant features. Suppose we run SVM classifier with features $X_1, X_2, \cdots, X_n$ and then add a irrelevant feature X_{n+1} that cannot help increase the margin. How will SVM classifier automatically ignore this feature? Justify your answer.