CIS519: Applied Machine Learning	**Spring 2018**
Quiz 9	
Handed Out: April 27, 2018	*Due: April 30,* 2018

1. What is the number of parameters needed to represent a Naive Bayes classifier with n Boolean variables and a Boolean label ?
   Options:

   (a) $2n + 1$

   (b) $n + 1$

   (c) $2n$

   (d) n

   Ans: (a)

2. Consider the following two graphic models 1a and 1b.

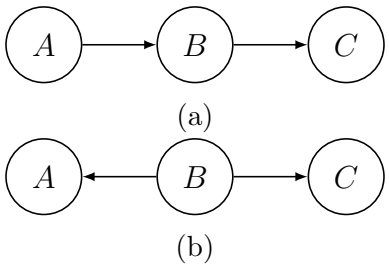

   Figure 1

   The joint probability distribution in 1a and 1b is equal to:

   (a) $P(A, B, C) = P(A)P(B)P(C)$ for both 1a and 1b

   (b) $P(A, B, C) = P(A)P(B|A)P(C)$ for 1a and $P(A, B, C) = P(B)P(B|A)P(B|C)$ for 1b

   (c) $P(A, B, C) = P(A)P(B|A)P(C)$ for 1a and $P(A, B, C) = P(B)P(A|B)P(C|B)$ for 1b

   (d) $P(A, B, C) = P(A)P(B|A)P(C|B)$ for both 1a and 1b

   Ans: (d)

3. If we train a Naive Bayes classifier using infinite training data that satisfies all of its modeling assumptions (e.g., conditional independence), then in general, what can we say about the training error (error in training data) and test error (error in held-out test data)?

   Options:

(a) It may not achieve either zero training error or zero test error

(b) It will always achieve zero training error and zero test error.

(c) It will always achieve zero training error but may not achieve zero test error.

(d) It may not achieve zero training error but will always achieve zero test error.

Ans: (a)

4. Which of the following is able to approximate any continuous function to an arbitrary accuracy?

Options:

(a) A two-layer neural network (input layer, output layer) using a linear activation function.

(b) A two-layer neural network (input layer, output layer) using a non-linear activation function.

(c) A three-layer neural network (input layer, hidden layer, output layer) using a linear activation function.

(d) A three-layer neural network (input layer, hidden layer, output layer) using a non-linear activation function.

Ans: (d)

5. What is the definition of the maximal margin in the context of SVM?

(a) $\max_{w;||w||=1} \min_{(x,y)\in S} |yW^T x|$

(b) $\min_{w;||w||=1} \max_{(x,y)\in S} |yW^T x|$

(c) $\min_{w;||w||=1} \max_{(x,y)\in S} yW^T x$

(d) $\max_{w;||w||=1} \min_{(x,y)\in S} yW^T x$

Ans: (a)

6. The use of sigmoid functions makes back-propagation possible because it is continuous and differentiable. Besides enabling back-propagation, the sigmoid function also makes neural network a:

(a) non-linear classifier

(b) linear classifier

Ans: (a)

7. One marble jar has several different colored marbles inside of it. It has 1 red, 2 green, 4 blue, and 8 yellow marbles. All the marbles are the same size and shape. If Peter takes out a marble from the jar without looking, what is the probability that he will NOT choose a yellow marble.

(a) $\frac{7}{15}$

(b) $\frac{8}{15}$

(c) $\frac{7}{8}$

(d) $\frac{5}{8}$

Ans: (a)

8. Let t_k and o_k be target value and output value of output unit k of a neural network. Let R be the learning rate. What is the delta rule (weight updating rule) of output unit in the backpropagation algorithm for neuron i's j-th weight $w_{i,j}$?

   (a) $\nabla w_{i,j} = R(t_i - o_i)o_i(1 - o_i)x_{i,j}$

   (b) $\nabla w_{i,j} = R(t_j - o_j)o_j(1 - o_j)x_{i,j}$

   (c) $\nabla w_{i,j} = R(t_j - o_j)t_j(1 - o_j)x_{i,j}$

   (d) $\nabla w_{i,j} = R(t_i - o_i)t_i(1 - o_i)x_{i,j}$

   Ans: (b)

9. In the AdaBoost algorithm, if the final hypothesis makes no mistakes on the training data, which of the following is correct?

   (a) Additional rounds of training can help reduce the errors made on unseen data.

   (b) The individual weak learners also make zero error on the training data.

   (c) Additional rounds of training always leads to worse performance on unseen data.

   Ans: (a)

10. Which of the following statements about the K-means algorithm are correct?

    (a) The K-means algorithm is sensitive to outliers.

    (b) The centroids in the K-means algorithm may not be any observed data points.

    (c) For different initializations, the K-means algorithm will definitely give the same clustering results.

    (d) The K-means algorithm can detect non-convex clusters.

    Ans: (a), (b)