Machine Learning, test 4

Name: .....

Student Number: .....

Complete the statements about training/testing

We use train set ...(1).....	A. to classify the input in training set
We use test set ...(2)....	B. to keep one for testing and the rest for training
We use 0.7-0.3 coin ...(3)....	C. to create the model and test if the model fits the data
We divide the dataset into folds ...(4)....	D. to split the dataset into training set and testing set respectively.
	D. to test if the model can handle new input well

1C 2E 3D 4B

Complete the following statements about SVM

To create linear svm ...(5).....	A. we use linear regression
To create non-linear svm ...(6)....	B. we use kernel function or transformation function
To create soft-margin svm ...(7)....	C. we use the slack variables
	D. we use dot product between two vector as kernel function

5D 6B 7C

Circle T or F for each statement below

8. Support vectors are near the decision boundary (confusion arear)	T	F
9. We can quickly find support vectors by sampling methods	T	F
10. After learning with SMO, we will have non-zero alpha coefficients for support vectors	T	F

8T 9F 10T

Complete the statements about the combination methods

Boosting is ...(11).....	A. to divide the sample space into homogeneous regions for minimizing the prediction error.
Bagging is ...(12).....	B. to combine trees, which are created by removing both rows and columns of the training data table
Random forest is ...(13).....	C. to combine trees from different bootstrapped datasets
Classification tree is ...(14).....	D. to combine trees created derived datasets via sample without replacement
	E. to create the linear combination of classifiers by emphasizing on misclassified samples.

11E 12C 13B 14A

Complete the following mixture models

You are given w1=3 and w2=7 and p1(x) = U(1, 3) and p2(x)=U(2, 5) to create the mixture model. Please complete the following statements about the model.

15. The model will be p(x) = …**0.3**…. * p1(x) + …**0.7**…… * p2(x) assuming that w1/p1 and w2/p2 are in pairs.

16. The value p(3) of the model will be …**0.3/(3-1) + 0.7/(5-2)**…….

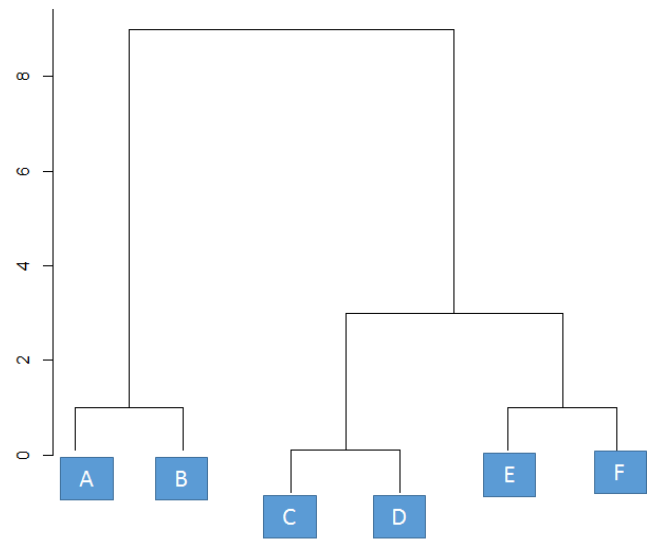
17. The value p(0) will be ……**0**…. and that of p(6) will be ……**0**…….

Select K if the statement is for kmeans, E if it is for EM, or N if it is for neither of them

18. Each sample can belong to one center	K	E	N
19. We need to initialize the centers after we have the labels	K	E	N
20. We can soft-assign a sample to centers	K	E	N

18K 19N 20E

Hierarchical clustering



21. When we build the tree from the dataset (1, 2, 7, 7.1, 9, 10) the label will be

A	B	C	D	E	F
1	**2**	**3**	**4**	**5**	**6**

22. And when we cut the tree into 2 classes the labels will be

1	2	7	7.1	9	10
1	**1**	**2**	**2**	**2**	**2**

Circle H for hierarchical, S for spectral, or K for kmeans for each statement below

23. We need to use eigen-decomposition to find the clusters	H	S	K
24. We can use group two small clusters into one larger cluster.	H	S	K
25. We start by initializing labels or centers	H	S	K
26. We can use it to build a "readable structure" for human to understand the dataset	H	S	K
27. We can fix the number of iteration in the learning/building process	H	S	K

23S 24H 25K 26H 27K

Complete the following statements

The dataset D is x1 = (100, 0, 0) x2 = (101, 0, 0) x3 = (102, 0, 0) x4 = (0, 200, 0) x5 = (0, 201, 0) x6 = (0, 202, 0) x7 = (0, 0, 300) x8 = (0, 0, 301) x9 = (0, 0, 302)	The first vocab V1 is m1 = (100, 0, 0) m2 = (0, 200, 0) m3 = (0, 0, 300) The first vocab V2 is m1 = (101, 0, 1) m2 = (1, 201, 0) m3 = (0, 1, 301)	28. The L1 error of V1 subject to D is ……**9**…… 29. The L1 error of V2 subject to D is …**15**…… 30. We can conclude that A. The vocab V1 is better for D B. The vocab V2 is better for D C. we can choose either of them for D, they are about the same.

30A

Uniform (U) or Importance (I) or rejection (R) or Box-Muller (B) sampling method?

31. We can draw N(0, 1) samples from this method	U	I	R	B
32. We need to compare the value f(x) where x is drawn from q(x) with another value drawn from the uniform distribution	U	I	R	B
33. We can draw any number within an interval (a, b) with this method	U	I	R	B
34. We can add twelve independent samples of N(0, 1) and minus six to have the similar result of this method	U	I	R	B
35. We don't need reject the sample but we can retain all samples with the associated value f(x)/q(x)	U	I	R	B

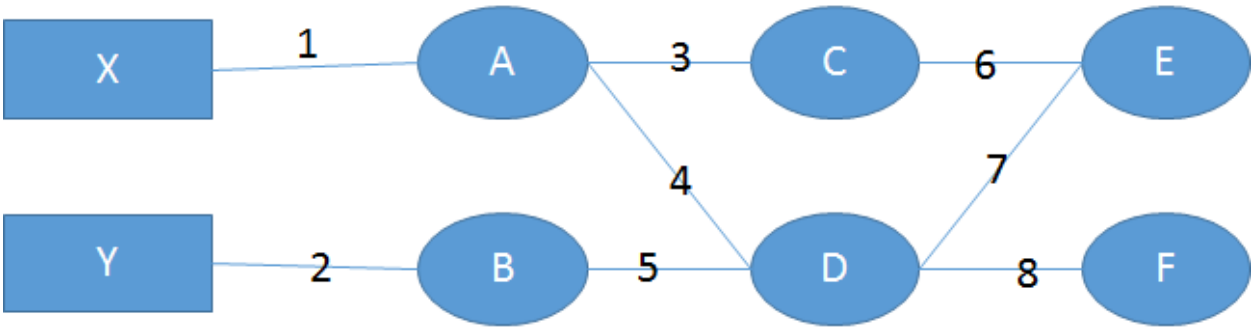
31B 32R 33U 34B 35I

Deep directed network (D) or deep Boltzmann machine (B) or deep belief network (F)

36. It has both arrow and non-arrow links	D	B	F
37. We can build it by just Bernoulli distributions and logistic functions	D	B	F
38. We can write the model using one exp function and one partition function	D	B	F

36F 37D 38B

Complete the equation for the following model



39. The first layer will be ……**AX+2BY**…….

40. The second layer will be ……**3AC+4AD+5BD**…….

41. The third layer will be ……**6CE+7DE+8DF**……..