

# CS 361 Machine Learning Exam 4 (January-May Session, 2021) 19th April 2021

Total Marks.: 40  
Time: 60 Minutes

Attempt all questions

...

Points: **34/40**



1

What happens when we get features in a lower dimension after applying PCA?  
(0/1 Point)

- ☐ Features must carry all the information in the original data
- ☒ Features will lose interpretability ✓
- ☐ Pairwise distances between original input and reconstructed input using principal components are minimized ✓
- ☐ None of the above

2

Which of the following is true about reinforcement learning (RL)?  
(1/1 Point)

- ☒ An agent gets reward or incurs a penalty according to an action ✓
- ☒ RL algorithms can update themselves incrementally with each newly observed example ✓
- ☒ To prioritize the rewards in the distant future, the discount factor should be close to 1 ✓
- ☒ A recommendation system is an application of RL ✓



3

In traditional neural network training,  
(0/1 Point)

- ☒ Weights can be regularized but not bias using a same regularization ✓
- ☐ Both Weights and Biases can be regularized ✓
- ☐ Both Weights and Biases can be regularized using a same regularization
- ☐ None of the above

4

In PCA, the principal component (say  $w_1$ ) will  
(1/1 Point)

- ☒ be the eigenvector of a covariance matrix of the input data with largest eigenvalue ✓
- ☐ be such that after projection of the data on to  $w_1$ , the difference between data sample points becomes least apparent
- ☒ have  $(w_1)'w_2 = 0$ , where  $(w_1)'$  is the transpose of  $w_1$  and  $w_2$  is the second principal component ✓

- ☐ need to use class labels at some point of time eventually

5

Which of the following statement(s) is/are TRUE?  
(1/1 Point)

- ☒ Principal component analysis (PCA) can be used for image compression. ✓
- ☐ Principal Component Analysis (PCA) finds the direction that maximize difference between two classes
- ☐ Linear Discriminant Analysis (LDA) find direction that maximize the variance in the data
- ☐ None of the Above



6

Considering the dimensionality reduction, which of the following is/are true?  
(0/1 Point)

- ☒ Pearson correlation can be used to identify variables with lower correlation.
- ☒ Eigen-decomposition is performed in PCA ✓
- ☐ Gaussian Discriminant Analysis (GDA) is a linear version of LDA.
- ☐ All of the above

7

You are using multilayer neural networks. You notice that the training error is going down and converges to a local minimum. Then when you test on the new data, the test error is abnormally high. What is probably going wrong and what would you do?  
(1/1 Point)

- ☒ The training data size is not large enough. Collect a larger training data and retrain it ✓

- ☒ Tune the learning rate and add regularization term to the objective function ✓
- ☐ Use the same training data but add a few more hidden layers
- ☒ Try k-fold cross validation ✓

8

Which of the following case(s) is/are NOT example(s) of Inductive Bias?  
(1/1 Point)

- ☐ The simplest consistent hypothesis about the target function is actually the best
- ☐ when forming a hypothesis, attempt to minimize the length of the description of the hypothesis
- ☐ when drawing boundary between two classes, attempt to maximize the width of the boundary
- ☒ None of the above ✓

9

Fill in the blanks: If we increase the k value in k-nearest neighbor, the model will \_\_\_\_\_ the bias and \_\_\_\_\_ the variance.  
(1/1 Point)

- ☒ Increase, Decrease ✓
- ☐ Decrease, Increase
- ☐ Decrease, Decrease
- ☐ Increase, Increase

10

Which of the following is True in the context of Neural Networks?  
(1/1 Point)

- ☒ Comparatively larger batches converge faster than smaller batches. ✓

- ☐ Comparatively larger batches converge slower than smaller batches.
- ☒ Larger batch sizes reduces the variance in the gradient estimation of SGD ✓
- ☐ Larger batch sizes induces the variance in the gradient estimation of SGD

11

In traditional AdaBoost, we use  
(1/1 Point)

- ☐ Linear loss function
- ☐ Ensembles of loss functions
- ☒ Exponential loss function ✓
- ☐ AdaBoost does not utilize any loss function

12

Which of the following data structure's operation can be aligned to that of Backpropagation?  
(1/1 Point)

- ☐ List
- ☐ Queue
- ☒ Stack ✓
- ☐ None of the Above

13

Suppose we have  $d$  features in our original dataset, and we use PCA to select the top  $k$  dimensions as the principal components. Which of the following is true?  
(1/1 Point)

- ☒ Higher k implies less regularization ✓
- ☐ If original dimensions are not correlated,  $k \ll d$
- ☐ Higher k means higher regularization
- ☒ A proper value of k can be decided by a scree plot of variance explained as a function of number of eigenvectors ✓

14

In an Unsupervised learning  
(1/1 Point)

- ☐ Specific output values are given
- ☒ Specific output values are not given ✓
- ☐ No specific Inputs are given
- ☐ Both inputs and outputs are given
- ☐ Neither inputs nor outputs are given

15

Which of the following shows deterministic nature in output?  
(1/1 Point)

- ☐ K-Means
- ☐ Neural Networks
- ☒ PCA ✓
- ☐ None of the above



16

Which of the following statement(s) is/are NOT true for a Restriction Bias?  
(0/1 Point)

- ☐ it restricts the hypothesis space to get simpler hypothesis
- ☐ "Linearity" may be imposed as a Restriction Bias
- ☐ Small decision tree over large decision tree ✓
- ☒ None of the above



17

For a classifier, the test error is related to the training error.  
(0/1 Point)

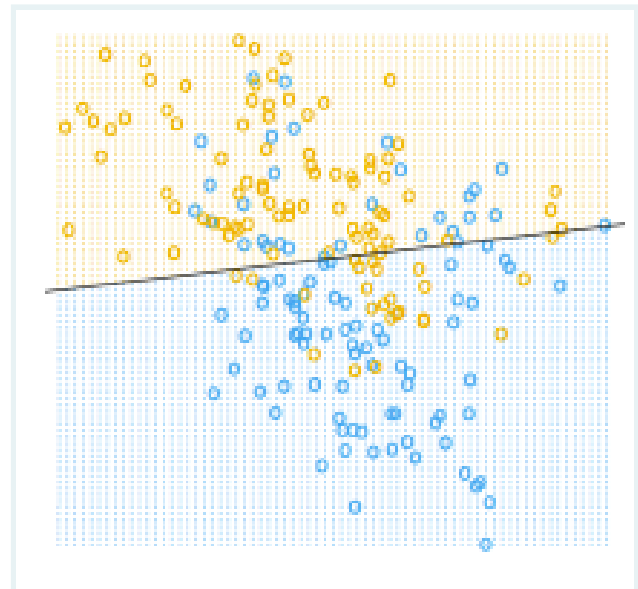
- ☐ TRUE ✓
- ☐ FALSE
- ☒ Can't say

18

Which is the following statement(s) is/are TRUE?  
(1/1 Point)

- ☒ In  $d$  dimensions,  $d-1$  dimensions will be orthogonal to the normal of any given hyper-plane ✓
- ☐ As distances between pairwise points become very small in high dimensional spaces, distances to hyperplanes become comparatively large
- ☒ most data points tend to be very close to these hyperplanes and it is often possible to perturb input slightly (and often imperceptibly) in order to change a classification outcome. ✓
- ☐ None of the above

19



Which of these classifiers could have generated such a decision boundary?  
(1/1 Point)

- ☐ k-NN (k=1)
- ☒ Linear SVM ✓
- ☒ Logistic regression ✓
- ☐ None of the above

20

Test error obtained using cross-validation is the \_\_\_\_ estimate of the generalization error.  
(1/1 Point)

- ☒ Biased ✓
- ☐ Unbiased
- ☐ Exact
- ☐ None of the above



21

Let  $(A - \lambda I)x = 0$  and  $B = A - \lambda I$ , then  $x$  will be an eigenvector of  $A$  if and only if

(1/1 Point)

- ☒  $\det(B)=0$  ✓
- ☒  $B$  does not have an inverse ✓
- ☐  $\det(B)$  is not equal to zero (0)
- ☐ None of the above

22

Suppose you are dealing with a classification problem with 8 classes. Then, at most how many discriminant vectors can be produced by LDA?

(1/1 Point)

- ☐ 16
- ☐ 8
- ☐ 9
- ☒ 7 ✓
- ☐ 64

23

Consider an image classification task using traditional neural networks, which of the following can be used as a computing function at the output layer?

(1/1 Point)

- ☒ tanH ✓
- ☐ ReLU

☒ Sigmoid ✓

☐ All of the above

24

A 5-input neuron has weights 4, 1, 7, 2, and 3. The transfer function is linear with the constant of proportionality being equal to 4. The inputs are 1, 10, 2, 5, and 20 respectively. The output will be  
(1/1 Point)

☐ 784

☒ 392 ✓

☐ 98

☐ 152

25

Which of the following is NOT TRUE for selecting base learners for an ensemble?  
(1/1 Point)

☐ Different learners can come from same algorithm with different hyper-parameters

☐ Different learners can come from different algorithms

☒ Different learners can come from same algorithm on same dataset but in different time ✓

☐ Different learners can come from same algorithm on different dataset

26

PCA is likely to show promising results when applied on a dataset if  
(1/1 Point)

☐ the data lies on a curved surface and not on a flat surface



- ☒ features are standardized first ✓
- ☒ there is a linear structure in the data ✓
- ☐ eigenvalues are roughly equal

27

With a soft-margin SVM, which samples will have non-zero slack variables?  
(1/1 Point)

- ☐ All samples lying on margin boundary
- ☐ All samples outside the margin
- ☒ All samples inside the margin ✓
- ☒ All misclassified samples ✓

28

Fill in the blanks: The VC dimension of a Perceptron is \_\_\_\_\_ the VC dimension of a linear SVM.  
(1/1 Point)

- ☐ larger than
- ☒ same as ✓
- ☐ smaller than
- ☐ not related at all to

29

While performing PCA, the prime objective is to  
(1/1 Point)

- ☐ pick a set of orthogonal dimensions from the original dimensions of the given dataset
- ☒ find a orthogonal basis such that the data has high variance along the basis

☒ components ✓

☐ find the components of the dataset

☐ estimate the number of dimensions

30

In an election,  $N$  candidates are competing against each other and people are voting for either of the candidates. Voters don't communicate with each other while casting their votes. Which of the following ensemble method works similar to the above-discussed election procedure?

(1/1 Point)

☒ Bagging ✓

☐ Boosting

☐ None of these

31

Consider the statement: Choosing different norms does not affect the decision boundary of  $k$ -NN ( $k=1$ ).

(1/1 Point)

☐ True

☒ False ✓

☐ Can't say

32

Say TRUE/FALSE: Ensemble of classifiers always be more accurate than any of its individual models.

(1/1 Point)

☐ TRUE

 FALSE ✓

33

Assume that we try to fit a linear and 8th-degree polynomial to data distribution coming from a cubic function, corrupted by standard Gaussian noise. Let M1 and M2 denote the models corresponding to the linear and 8th-degree polynomial. Then

(1/1 Point)

- ☐ Bias(M1)  $\leq$  Bias(M2), Variance(M1)  $\leq$  Variance(M2)
- ☐ Bias(M1)  $\leq$  Bias(M2), Variance(M1)  $\geq$  Variance(M2)
- ☒ Bias(M1)  $\geq$  Bias(M2), Variance(M1)  $\leq$  Variance(M2) ✓
- ☐ Bias(M1)  $\geq$  Bias(M2), Variance(M1)  $\geq$  Variance(M2)

34

Which of the following statement(s) is/are NOT TRUE?

(1/1 Point)

- ☐ Variance is the measure of the deviation from the mean for points in one dimension
- ☐ Covariance is the measure of how much each of the dimensions vary from the mean with respect to each other
- ☒ Variance is the statistical average of the every dimensions ✓
- ☐ None of the above

35

Which of the following statement(s) is/are TRUE for Bayesian Network?

(1/1 Point)

- ☒ Directed type of graphical model which represents conditional independence relationships between different variables in the domain ✓

- ☒ Helps to get tractable inference where all the variables are not connected (dependent) to each other ✓
- ☒ Interested to represents the dependence (or independence) between different variables in the domain ✓
- ☐ None of the above

36

In the context of ML, dimension reduction refers to  
(1/1 Point)

- ☐ Reducing number of trainable parameters
- ☒ Reducing number of random variables ✓
- ☐ Reducing number of hidden layer neurons
- ☐ None of the above

✗

37

Which of the following are true?  
(0/1 Point)

- ☒ To be able to apply LDA, the within-class-scatter-matrix should be singular.
- ☒ LDA is supervised whereas PCA is unsupervised ✓
- ☒ PCA maximizes the variance of the data only, whereas LDA maximizes the separation between different classes only
- ☒ Both LDA and PCA are linear transformation techniques ✓

38

Which of the following is/are true in the context of neural network training?  
(1/1 Point)

- ☐ Each weight is updated with the avg. value of gradients of all weights.
- ☒ A constant higher learning rate may result in local minima stuck ✓
- ☐ One should not apply non-linearity when number of hidden layers are  $< 3$ .
- ☐ Only a, b

39

For any feedforward network, we can always create an equivalent feedforward network with separate layers  
(1/1 Point)

- ☒ True ✓
- ☐ False
- ☐ Can't say

40

Find the eigenvalues of  
(1/1 Point)

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}$$

- ☐ -1, -1
- ☐ 1, 2
- ☒ -1, -2 ✓
- ☐ None of the above

