

CS 361 Mid Semester Exam (Jan - May 2022) 3rd March 2022 (3 PM - 5 PM)

Answer all questions. Time: 120 minutes Full Marks: 50

Points: 45/50

✓ **Correct** 1/1 Points

1

In neural networks, the common points about activation functions such as sigmoid, tanh, and ReLU is/are:

- ☐ Always output values between 0 and 1
- ☒ Help to learn nonlinear decision boundaries ✓
- ☐ Are applied only to the output units
- ☐ None

✓ **Correct 1/1 Points**

2

Which of the following statements is/are true?

- ☐ Considering an MLP and an image given as input, each neuron looks for patterns in different local regions of the input at a time
- ☐ The earlier layers of a neural network are typically computing more complex features of the input than the deeper layers.
- ☒ A neural network without non-linearity is equivalent to a linear regression model ✓
- ☒ For a neural network with ReLU activation function, sometimes neurons can stop responding to variations in the computed error ✓

✓ **Correct 1/1 Points**

3

What value would be in place of question marks?

- ☒ 0.9523, 0.0474, 0.0003 ✓
- ☐ 0.8, 0.1, 0.1
- ☐ 0.8523, 0.1474, 0.0003
- ☐ 0.9132, 0.1020, 0.0006

✓ **Correct** 1/1 Points

4

What if we use a learning rate that's too large?

- ☐ Network will converge.
- ☒ Network may not converge.
- ☐ Can't Say

✓

✓ **Correct** 1/1 Points

5

What steps can we take to prevent overfitting in a Neural Network?

- ☒ Data Augmentation
- ☒ Weight Sharing
- ☐ Remove Batch-Norm from first few layers of the network
- ☐ None of the above

✓

✓

✗ **Incorrect** 0/1 Points

6

Which of the following is/are used for modeling times series and sequential data?

- ☐ Decision graphs
- ☒ Dynamic Bayesian networks
- ☐ Gated recurrent units
- ☐ Context Aware CNN

✓

✓ **Correct** 1/1 Points

7

A 1-nearest neighbor classifier has _____ and _____ than a 3-nearest neighbor classifier.

- ☒ High Variance
- ☐ High Bias
- ☐ Low Variance
- ☒ Low Bias

✓

✓

✓ **Correct 1/1 Points**

8

Which of the following statement(s) is/are TRUE for kernel function of the SVM?

- ☒ Kernel function can be thought of a similarity measure between the input objects ✓
- ☐ Any similarity measure can be used as Kernel function
- ☒ Every semi-positive definite symmetric function is a kernel ✓
- ☐ None of the above

✓ **Correct 1/1 Points**

9

Bayesian networks allow compact specification of

- ☐ Conditional independence
- ☐ Belief
- ☐ Marginal distributions
- ☒ Joint probability distributions ✓
- ☐ Propositional logic statements

✓ **Correct 1/1 Points**

10

Match the solution of Overfitting to its disadvantage.

- | | |
|----------------------------|--|
| 1) K-fold Cross Validation | a) Leads to a high bias error |
| 2) Regularization | b) Model does not make use of all training data |
| 3) Early stopping | c) Constituting extra hyper-parameters at each layer |
| 4) Drop-out | d) Doesn't work well with sequential data layer |

☐ 1 - d, 2 - b, 3 - c, 4 - a

☒ 1 - d, 2 - a, 3 - b, 4 - c

✓

☐ 1 - c, 2 - d, 3 - a, 4 - b

☐ 1 - b, 2 - d, 3 - c, 4 - a

✓ **Correct 1/1 Points**

11

Why SVM's are in general believed to be faster (in terms of convergence rate) than other models for some problems?

☒ Due to its inclusion of the kernel trick in the convex optimization.

✓

☐ Due to its non-convex optimization approach.

☐ Due to its ability to perform when data samples are sufficiently not large.

☐ Due to the restricted number of allowed support vectors.

✓ **Correct 1/1 Points**

12

Consider a univariate linear regression model. Which of the following are true ?

- ☐ Changing the input variable by 1 unit always affects the output by 1 unit too.
- ☐ Since it is univariate, we need to estimate one coefficient for modelling the data.
- ☐ The decision boundary is a property of the data set given to us.

☒ Figure above shows two ways of computing residuals/offsets. We use the method as shown in the left side while calculating loss using mean squared error function. (Assume y-axis is the dependent variable)

✓

☐ None of the above.

✓ **Correct 1/1 Points**

13

Consider the given figure that shows a sample data consisting of two features X_1 and X_2 , and a class label Y . Assume that k-NN is used for memorizing this data and performing classification on some test points. But observing the data and labels, there could be some problems if k-NN memorizes this given data. What operations can be done so that k-NN can generally work well with such data too?

- ☒ Transform the data by column-wise subtracting the mean and dividing by the standard deviation.
- ☐ Plotting a graph of validation error-rate vs. k-value and get the best value of k
- ☐ Checking different distance metrics and deciding which performs the best.
- ☐ None of the above

✓

✓ **Correct 1/1 Points**

14

Which is/are CORRECT statement(s)?

- ☐ Always Sigmoid is a better choice than SoftMax for multiclass classification.
- ☒ Any function can be approximated to arbitrary accuracy by a neural network having two hidden layer. ✓
- ☐ Single layer neuron can solve some non-linearly separable problems.
- ☐ None of the above

✓ **Correct 1/1 Points**

15

Match the type of SVM kernel function to related mathematical representation.

- ☐ 1 - d, 2 - b, 3 - c, 4 - a
- ☐ 1 - b, 2 - d, 3 - c, 4 - a
- ☐ 1 - c, 2 - d, 3 - a, 4 - b
- ☒ 1 - d, 2 - a, 3 - b, 4 - c ✓

✓ **Correct** 1/1 Points

16

Match the related rows of given columns (Column 1: Structure of Perceptron, Column 2: Type of Decision Regions)

1) Single-Layer Perceptron a) Arbitrary (Complexity limited by number of nodes)

2) Two-Layer Perceptron b) Half Plane Bounded by Hyperplane

3) Three-Layer Perceptron c) Convex Open or Closed Regions

☐ 1 – c, 2 – a, 3 – b

☐ 1 – a, 2 – b, 3 – c

☐ 1 – b, 2 – a, 3 – c

☒ 1 – b, 2 – c, 3 – a

✓

✗ **Incorrect** 0/1 Points

17

Choose the correct option(s):

☒ log likelihood of an individual model is meaningless

✓

☒ Likelihood ratio test is an indicator of "goodness" of fit

✓

☐ Given a normally distributed data, likelihood and log likelihood functions peak at the same values of mean and variance of the data

☐ Given a normally distributed data $X = \{x_1, x_2, \dots, x_n\}$, the maximum likelihood estimate for where the center of the normal distribution should be can be computed as $(x_1 + x_2 + \dots + x_n)^2 / (2 \cdot n)$

☐ None of the above

✓ **Correct 1/1 Points**

18

Among the following steps in regression modelling, which one/more impact(s) the trade-off between overfitting and underfitting the most.

- ☒ The polynomial degree. ✓
- ☐ Whether to learn the weights by matrix inversion or gradient descent.
- ☐ The constant-term.
- ☐ None of the above.

✓ **Correct 1/1 Points**

19

Identify true/false statement:

Statement 1: In case of Overfitting, model have High Bias and High Variance.

Statement 2: Underfitting happens when a model captures only the underlying pattern of the data.

Statement 3: In case of underfitting model have High Bias and Low Variance.

Statement 4: Overfitting happens when our model captures the noise along with the underlying pattern in data.

- ☐ 1 – False, 2 – False, 3 – True, 4 - False
- ☒ 1 – False, 2 – False, 3 – True, 4 - True ✓
- ☐ 1 – True, 2 – True, 3 – False, 4 - True
- ☐ 1 – True, 2 – False, 3 – True, 4 - True

✗ **Incorrect** 0/1 Points

20

You are given a binary classification task (two classes - benign tumor and malignant tumor) using a linear SVM. Let x_p be vector denoting positive support vector and x_n be the vector denoting negative support vector, such that x_p lies on $wx+b = +1$ and x_n lies on $wx+b = -1$. Suppose the margin is $f(w)$ and we have two more functions : $g(z) = 1/z$ and $h(q) = (q^2)/2$. Which of the following hold true?

- ☒ The margin of this SVM is given by the projection of the vector addition between x_p and x_n ✗
- ☐ While solving the lagrangian duality, dimensionality of the solution gets reduced
- ☐ Say $f(w)$ becomes a decreasing function over the positive domain of real values. Then maximizing $f(w)$ is equivalent to maximizing $g(f(w))$
- ☒ Minimizing $g(f(w))$ is equivalent to minimizing $h(g(f(w)))$ ✓

✓ **Correct** 1/1 Points

21

In a supervised classification problem, which statement(s) is/are TRUE

- ☐ Most probable hypothesis always ensures most probable classification
- ☒ The most probable classification of the new instance is obtained by combining the predictions of all hypotheses, weighted by their posterior probabilities. ✓
- ☐ The hypothesis functions in a hypothesis class can not be equiprobable.
- ☐ None of above.

✓ **Correct 1/1 Points**

22

Find odd man out

- ☐ Sigmoid
- ☒ KKT conditions
- ☐ Radial Basis Function
- ☐ Polynomial of power p

✓

✓ **Correct 1/1 Points**

23

Which statement(s) is/are TRUE in the case of neural network?

- ☒ The single layer neural network has limited representation power.
- ☐ To solve any problem neural network with at least three hidden layers is required.
- ☐ A neural network with a single layer cannot realize every Boolean function.
- ☒ The input layer of neural network does not possess any activation function.

✓

✓

✓ **Correct** 1/1 Points

24

You are training multilayer neural networks using Batch Gradient Descent (BGD) and least squares method as the cost function. 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?

- ☒ The training data size is not large enough. Collect a larger training data and retrain it. ✓
- ☐ Use the same training data but add a few more hidden layers
- ☐ Do Stochastic Gradient Descent instead of Batch Gradient Descent
- ☐ Try a different cost function
- ☐ None of the above

✓ **Correct** 1/1 Points

25

What will give best results for a Bayesian classification when some features are missing?

- ☐ We assuming the missing values as the mean of all values.
- ☐ We ignore the missing features.
- ☒ We integrate the posteriors probabilities over the missing features. ✓
- ☐ Drop the features completely.

✓ **Correct 1/1 Points**

26

Which of these classifiers could have generated such a decision boundary?

- ☐ k-NN (k=1)
- ☒ Linear SVM ✓
- ☒ Single Layer Perceptron ✓
- ☐ SVM with Gaussian kernel
- ☐ None of the above

✓ **Correct 1/1 Points**

27

Assertion (P): Consider a Linear Regression model learned for a dataset with two features - height and weight. The distribution of the data can be seen in the figure. Using Stochastic Gradient Descent (SGD) in such a case instead of Batch Gradient Descent (BGD) is a good idea.

Reason (R): In each iteration of BGD, gradients will be recomputed for similar training examples before each update of model parameters, which may not carry much predictive value.

- ☒ Both P and R are true and R is the correct explanation of P. ✓
- ☐ Both P and R are true but R is not the correct explanation of P.
- ☐ Both P and R are false
- ☐ P is false but R is true
- ☐ P is true but R is false

✓ **Correct 1/1 Points**

28

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 8 degree polynomial. Then:

☐ 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)

✓ **Correct 1/1 Points**

29

In case of following soft margin SVM classifier, the role of variable C is/are

☐ It controls the rate of convergence.

☒ It controls the width of the margin. ✓

☒ It controls overfitting. ✓

☐ It helps dual formation

✗ **Incorrect** 0/1 Points

30

Domain(s) where one can use neural network?

- | | |
|---|---|
| <input checked="" type="checkbox"/> Autopilot enhancements | ✓ |
| <input checked="" type="checkbox"/> Deterministic finite automata | ✗ |
| <input checked="" type="checkbox"/> Speech synthesis | ✓ |
| <input checked="" type="checkbox"/> Trajectory control | ✓ |

✓ **Correct** 1/1 Points

31

Choosing different norms (distance function) does not affect the decision boundary of k-NN ($k=1$).

- | | |
|---|---|
| <input type="checkbox"/> True | |
| <input checked="" type="checkbox"/> False | ✓ |
| <input type="checkbox"/> Can't Say | |

✓ **Correct** 1/1 Points

32

If all the weights are set to zero instead of random initializations in NN for a classification task, what can be an expected behavior?

- ☐ No problem. The NN will train properly.
- ☒ The NN will train. However, all the neurons will end up recognizing the same thing. ✓
- ☐ The NN will not train.
- ☐ None of these.

✓ **Correct** 1/1 Points

33

In the Naive Bayes algorithm, suppose that prior for class w_1 is greater than class w_2 , would the decision boundary shift towards the region R_1 (region for deciding w_1) or towards region R_2 (region for deciding w_2)?

- ☐ towards region R_1 .
- ☒ towards region R_2 . ✓
- ☐ No shift in decision boundary.
- ☐ It depends on the exact value of priors.

✓ **Correct 1/1 Points**

34

Let x, y, z be three variables taking binary values. Define $f(x, y, z) = x'y'z + x'yz' + xy'z' + xyz$, then the data set is linearly separable.

☐ True

☒ False

✓

☐ If we set $x = 0$, $y = 0$, and $z = 0$, then True

☐ Can't Say.

✓ **Correct 1/1 Points**

35

What is the consequence between a node and its predecessors while creating Bayesian network?

☐ Functionally dependent

☐ Dependent

☒ Conditionally independent

✓

☐ Both Conditionally dependent & Dependent

✓ **Correct** 1/1 Points

36

For a k-NN classifier for a d-dimensional dataset having n data points, which of the following statement(s) is/are TRUE?

- ☐ If n approaches infinity, k-NN classifier starts random guessing.
- ☒ As $d \gg 0$, k-NN assumption (i.e. similar points share similar labels) breaks down ✓
- ☐ k-NN can't perform classification task well in supervised setting
- ☐ None of the above

✓ **Correct** 1/1 Points

37

With a soft-margin SVM, which samples will have non-zero slack variables ?

- ☐ All samples lying on margin boundary
- ☐ All samples outside the margin
- ☒ All samples inside the margin ✓
- ☒ All misclassified samples ✓
- ☐ Can't say

✓ **Correct** 1/1 Points

38

Which of the following is/are true about Maximum Likelihood Estimate (MLE)?

- ☒ MLE does not necessarily exist. ✓
- ☐ MLE always exists.
- ☒ If exists/exist, it/they may not be unique. ✓
- ☐ If exists/exist, it/they must be unique.

✗ **Incorrect** 0/1 Points

39

Regarding performance of the supervised classifier, which of the following statement(s) is/are FALSE

- ☒ There is a relation between training and testing error. ✗
- ☐ Low training error does not guarantee low testing error.
- ☐ High generalization error always suggests high training error
- ☐ None of the above

✓ **Correct 1/1 Points**

40

We can get multiple local optimum solutions if we solve a linear regression problem by minimizing the sum of squared errors using gradient descent.

☐ True

☒ False

✓

☐ Depends on the momentum term of the SGD

☐ Can't say

✓ **Correct 1/1 Points**

41

Given an SVM, which of the following are true?

☒ KKT conditions state that solution for dual problem is same as solution to the primal problem

✓

☒ We solve the dual instead of the primal problem because doing so helps the optimization problem as we have to compute only inner products between training samples

✓

☐ Using KKT conditions it can be implied that the optimal α_i (i.e. lagrange multiplier associated with i -th training example) will be zero for all support vectors, helping us compute inner products for only these support vectors now

☐ None of the above

✓ **Correct 1/1 Points**

42

What is the use of regularization parameter while performing a regularized linear regression (RLR)?

- ☒ Until some point, increasing it reduces the variance of the model significantly without significant addition of bias to the model ✓
- ☐ It reduces the bias in the model and hence reduces overfitting
- ☐ Controls the trade-off between the need for the model to fit the training set well and also have a large number of model parameters
- ☐ Helps to find the exact decision boundary regardless of its complexity

Consider a self-driving car that learns an RLR model X that gives the best driving performance based on 10 attributes – e.g., road curvature, steering angle and speed. After a month, you get data about 15 more attributes like weather, driver experience, and the car model - and incorporate them into X. Suppose the re-trained RLR model is Y. A high regularization parameter increases the inability of Y to capture the true relationship between the 25 attributes in the dataset.

- ☒ ✓

✓ **Correct** 1/1 Points

43

The neural network given above takes two binary valued inputs x_1, x_2 and the activation function is the binary threshold function ($h(x)=1$ if $x>0$, otherwise 0). Which of the following logical functions does it compute?

☐ AND

☐ OR

☒ NAND

☐ NOR

☐ XOR

✓

✓ **Correct** 1/1 Points

44

Assertion (P): The path taken by Stochastic Gradient Descent (SGD) towards the minima always has low variance and faster convergence as compared to Batch gradient descent.

Reason (R): In each iteration of SGD, random samples are picked based on which gradients are calculated.

☐ Both P and R are true and R is the correct explanation of P.

☐ Both P and R are true but R is not the correct explanation of P.

☐ Both P and R are false.

☒ P is false but R is true.

☐ P is true but R is false.

✓

✓ **Correct** 1/1 Points

45

Identify the true/false statement regarding gradient decent optimization technique:

Statement 1: A gradient measures the degree of change of a variable in response to the changes of another variable.

Statement 2: Gradient Descent is a concave function whose output is the partial derivative of a set of parameters of its inputs.

- ☐ Both statements are true
- ☐ Statement 1 is false and 2 is true
- ☒ Statement 1 is true and 2 is false
- ☐ Both statements are false

✓

✓ **Correct 1/1 Points**

46

You, a data scientist, are approached by Mr. Biju from the Kerala State Archives Department, with a letter written in Malayalam that dates back to 1906. He tells you that the letter contains information regarding the last painting of Raja Ravi Varma – one of India's foremost artists. The painting remains lost for a century - but this letter, written by Raja himself, contains encrypted messages in Malayalam that might reveal the painting's location. Before messages can be decrypted by computers, Mr. Biju wants you to build a model for handwriting recognition. You decide to use an SVM for the same. Which of the following hold true?

- ☒ Since you have 44 basic characters in the language, you can use one-vs-all method using 44 binary classifiers ✓
- ☒ Soft margin can be employed to improve generalizability of SVM ✓
- ☒ On adopting a hard-margin SVM classifier, gradient descent can guarantee optimal solution ✓
- ☒ Effectiveness of SVM depends upon kernel function selected for the task at hand ✓

✓ **Correct 1/1 Points**

47

Which of the following gives non-linearity to a neural network?

- ☐ Stochastic Gradient Descent
- ☒ Rectified Linear Unit ✓
- ☐ Convolution operator
- ☐ None of the above

✓ **Correct 1/1 Points**

48

The robotic arm will be able to paint every corner in the automotive parts while minimizing the quantity of paint wasted in the process. Which learning technique is used in this problem?

- ☐ Supervised Learning.
- ☐ Unsupervised Learning.
- ☒ Reinforcement Learning. ✓
- ☐ Both (A) and (B).

✓ **Correct 1/1 Points**

49

Consider the following problem of linear SVM, which of the following statement(s) is/are CORRECT?

- ☐ This is a linear optimization problem subject to convex quadratic constraints.
- ☒ This is a convex quadratic optimization problem subject to linear constraints. ✓
- ☒ This can be solved using Quadratic Programming (QP) ✓
- ☐ None of the above.

✓ **Correct 1/1 Points**

50

What are some common ways to tackle overfitting?

- ☒ Cross-validation ✓
- ☒ Early stopping ✓
- ☐ Train with less data
- ☒ Regularization ✓

This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

Powered by Microsoft Forms | [Privacy and cookies](#) | [Terms of use](#)