CS 361 Machine Learning Exam 2 (January-May Session, 2021) 1st March 2021

Total No.: 20 Attempt all questions Time: 30 Minutes

Points: 13/20

1

On training a predictive model using some learning algorithm, it is apprehended that the model is underfitting. Which of the following could be the nature of predictive errors ?
(1/1 Point)

- [ ] Low bias and High variance
- [ ] Low bias and low variance
- [x] High bias and low variance ✓
- [ ] High bias and high variance

2

Why do we need the soft margin formulation in an SVM classifier?
(0/1 Point)

- [ ] To improve the generalization ability of SVM ✓
- [x] To be able to perform classification of data which is not linearly separable
- [ ] To avoid misclassification as much as possible while keeping a small margin
- [x] To allow a certain trade-off between number of misclassifications and margin width ✓

3

Which of the following hold true for SVMs?
(0/1 Point)

- [ ] Moving only the support vectors around affects the separating hyperplane as well ✓
- [x] On adopting a hard-margin SVM classifier, gradient descent can guarantee optimal solution ✓
- [x] On adopting a binary soft-margin SVM classifier, if any datapoint is found on the wrong side of margin, it incurs a penalty of 'p' units, otherwise no (0 units) penalty on it.
- [x] Sensitive to noise ✓

4

What are some common ways to tackle overfitting?
(1/1 Point)

- [ ] Putting non linearity
- [ ] Setting bias value to zero
- [x] Train with more data ✓
- [ ] kernel trick

5

What if we use a learning rate that's too large?
(1/1 Point)

- [ ] Network will converge
- [x] Network may not converge ✓
- [ ] Can't Say

✗

6

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

- [ ] In a single layer perceptron, inputs are limited to boolean values
- [x] In backpropagation, we should start with a small learning parameter and slowly increase it during the learning process
- [x] A single layer perceptron can be used to implement only linearly separable functions. ✓
- [ ] NAND, NOR and XOR can be implemented via a single layer perceptron

7

Which of these can be successfully implemented via a standard feed-forward neural network?
(1/1 Point)

- [ ] We need to learn shared features across different positions of text input.
- [x] 400,000 text reviews from different customers were collected for Amazon products. Each review is also labeled either positive or negative. The task is to train your model and predict on test reviews to label them as either positive or negative. ✓
- [ ] Inputs and outputs are of different lengths in different training examples.

None of the above

8

We can get multiple local optimum solutions if we solve a linear regression problem by minimizing the sum of squared errors using gradient descent.
(1/1 Point)

- [x] False ✓
- [ ] True
- [ ] Depends on the momentum term of the SGD
- [ ] Can't say

9

Question
(1/1 Point)

When solving for a hyperplane specified by $w^T x + b = 0$, one can always set i.e., $w^T x_1 + b = -1$, and $w^T x_2 + b = 1$?

- [ ] Always True
- [x] True, but as long as it is separable. ✓
- [ ] Always False
- [ ] False, only when x1 > x2. True otherwise.

10

Why SVM's are in general believed to be faster (in terms of convergence rate) than other models for some problems?
(0/1 Point)

- [x] Due to its inclusion of the kernel trick in the convex optimization ✓

- [ ] Due to its non-convex optimization approach.
- [x] Due to its ability to perform when data samples are sufficiently not large
- [ ] Due to the restricted number of allowed support vectors

11

What will happen if we set all the weights to zero instead of random weight initializations in NN for a classification task?
(0/1 Point)

- [ ] There won't be any problem and the NN will train properly
- [ ] The NN will train but all the neurons will end up recognizing the same thing ✓
- [x] The NN will not train as there is no net gradient change
- [ ] None of these

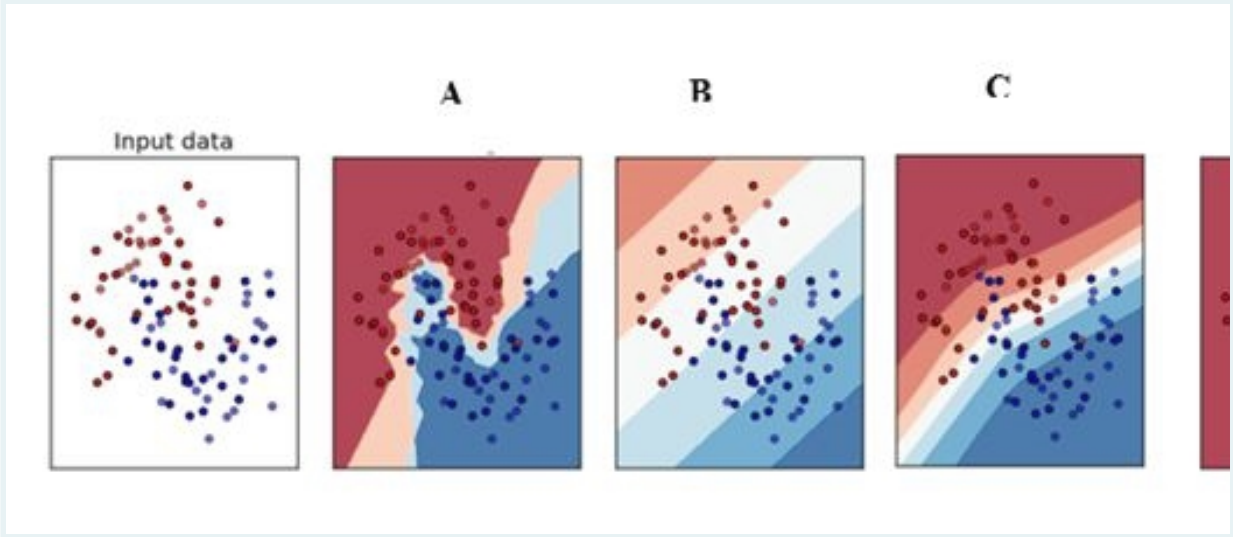
12

Find odd man out
(1/1 Point)

- [ ] Radial Basis Function
- [ ] Polynomial of power p
- [ ] Sigmoid
- [x] KKT conditions ✓

13

A 5-layer feed-forward neural network is built to solve a multi-class classification problem. Which of the following is a decision boundary generated by the network?
(0/1 Point)



- [ ] D ✓
- [ ] B ✓
- [x] A ✓
- [ ] C ✓
- [ ] None of these

14

If the given data samples are not linearly separable, which of the following classifier can't able to classify?
(1/1 Point)

- [ ] Multi-layer perceptron (MLP)
- [ ] Logistic regression
- [x] Single Layer Perceptron ✓
- [ ] Support Vector Machine (SVM)

15

Which of the following are the drawbacks of backpropagation?
(1/1 Point)

- [x] Can get stuck in local minima
- [x] Slow convergence possible
- [x] Scaling problem
- [ ] Cannot help to learn weights and biases for too much complex functions.

16

Effectiveness of SVM depends upon
(1/1 Point)

- [x] Selection of kernel
- [x] Soft margin parameter C
- [x] Kernel parameters
- [ ] Non-support vectors

17

A multilayer neural network where the neurons operate in the linear region can be approximated to a single layer network.
(1/1 Point)

- [x] True
- [ ] False
- [ ] Can't say

18

In neural networks, activation functions (sigmoid, Tanh, and ReLU) (1/1 Point)

- [ ] Speed up the gradient calculation in backpropagation, as compared to linear units
- [ ] are applied only to the output units
- [x] help to learn nonlinear decision boundaries ✓
- [ ] always output values between 0 and 1

19

While training a neural network, it is observed that the loss decreases heavily during the first few epochs. But then it becomes stagnant over a large number of epochs, and then suddenly starts decreasing rapidly. To solve this issue, a helper function is written - solve_problem_X(). Which of the following approaches could be the contents of solve_problem_X()? (0/1 Point)

- [ ] Increase the number of parameters to avoid falling into local minima
- [x] Decay the learning rate linearly until a fixed number of epochs and then keep it constant ✓
- [x] If validation loss does not improve for a fixed number of epochs, reduce the learning rate by a factor of 2-10 ✓
- [ ] To the weight update process, add an exponentially weighted average of the previous gradients ✓
- [ ] None of these

20

Consider the following two statements:

Statement 1 (S1): If hypothesis is richer, overfitting is not likely.

Statement 2 (S2): If feature space is larger, overfitting is more likely.
(1/1 Point)

- [ ] S1 is true but S2 is false
- [x] S2 is true but S1 is false ✓
- [ ] Both are true
- [ ] Both are false