

## Quiz 1 - Results



### Attempt 1 of 1

Written Feb 4, 2024 7:37 PM - Feb 4, 2024 7:55 PM

Released Feb 9, 2024 5:30 PM

Attempt Score **22.75 / 25 - A-**

Overall Grade (Highest Attempt) **22.75 / 25 - A-**

#### Question 1

1.5 / 1.5 points

As the model complexity increases, what is typically seen (Select all that apply)

- ✓ ☐ The testing accuracy decreases and then increases
- ✓ ☐ The training accuracy increases
- ✓ ☐ The training accuracy changes but testing does not
- ✓ ☐ The testing accuracy increases and then decreases

#### Question 2

1.5 / 1.5 points

Which of the following are part of the Model Development process? (Select all that apply)

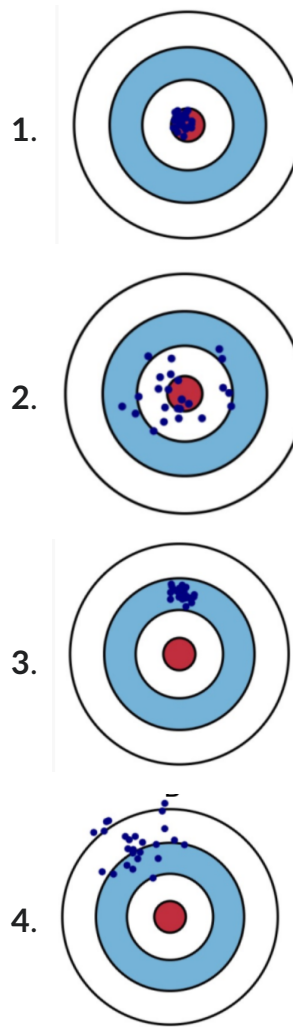
- ✓ ☐ Model Evaluation
- ✓ ☐ Model Deployment
- ✓ ☐ Hyperparameter Tuning
- ✓ ☐ Data Preprocessing

#### Question 3

1.5 / 1.5 points

Match each image with the bias-variance combination it depicts.

- ✓ \_\_3\_\_ Low Variance-High Bias
- ✓ \_\_1\_\_ Low Variance-Low Bias
- ✓ \_\_2\_\_ High Variance-Low Bias
- ✓ \_\_4\_\_ High Variance-High Bias



#### Question 4

1.5 / 1.5 points

Select all statements that are true

- ✓ ☐ Logistic regression predicts a continuous dependent variable.
- ✓ ☐ Logistic regression predicts a categorical/discrete dependent variable.
- ✓ ☐ Linear regression predicts a categorical/discrete dependent variable.
- ✓ ☐ Linear regression predicts a continuous dependent variable.

#### Question 5

1.5 / 1.5 points

Which of the following can be considered a hyperparameter? (multiple answers)

- ✓ ☐ The coefficients (or weights) of linear and logistic regression models
- ✓ ☐ Train-test split ratio
- ✓ ☐ Number of iterations a model is trained for
- ✓ ☐ Number of neighbors (k) in K-Nearest Neighbors algorithm

#### Question 6

0.75 / 1.5 points

In holdout validation a dataset is split into a train and test set through random subsampling. Randomly 3/4 of the samples in the dataset are selected for the training set and remaining 1/4 of the samples are taken as the test set.

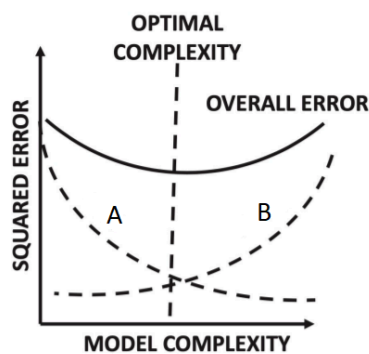
Select all that is true.

- ☒ The train and test sets may be imbalanced and the proportion of a class in each of the sets may be different from the original dataset.
- ☒ The train and test will always have the same distribution as the dataset was split randomly.
- ☒ Since we are holding out test set from the training process, the performance estimate is not biased.
- ☒ If the dataset is big enough random subsampling should be preferred even when the dataset is highly imbalanced.

### Question 7

0.75 / 1.5 points

Identify the labels A and B.



Select all that apply.

- ☒ A - Test Error, B - Training Error
- ☒ A - Variance, B - Bias
- ☒ A - Training Error, B - Test Error
- ☒ A - Bias, B - Variance

### Question 8

1.5 / 1.5 points

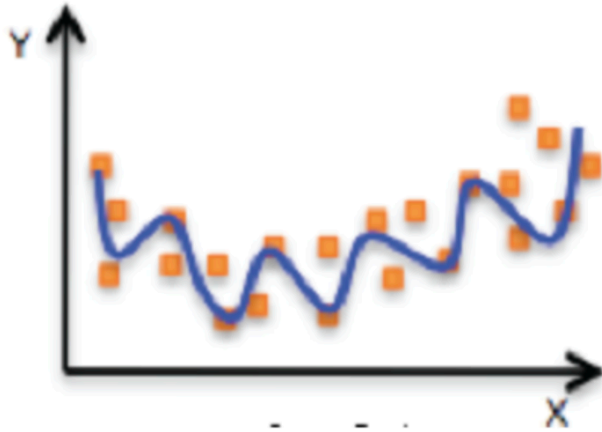
Select all that is true about bias and variance in machine learning. (Select all that apply)

- ☒ Training a complex model with a large amount of training data should help in reducing variance.
- ☒ When a model has a high bias, we can lower it by training with a large amount of data.
- ☒ Bias is the error caused due to simplistic assumptions in the model.

- ✓ ☐ When the variance is high, increasing the model capacity alone will help in reducing the variance.

**Question 9****0.75 / 0.75 points**

What does the following graph depict?



- ✓ ☐ Overfitting
- ☐ Linear Fit
- ☐ Underfitting
- ☐ Inconclusive

**Question 10****1.5 / 1.5 points**

Which parameter tells you how close the regression line is to a set of points? (More than one might apply)

- ✓ ☐ Mean squared Error
- ✓ ☐ Root mean squared error
- ✓ ☐ Bias
- ✓ ☐ Variance

**Question 11****0.75 / 0.75 points**

What type of learning do Logistic and Linear Regression fall under?

- ✓ ☐ Supervised
- ☐ Semi-Supervised
- ☐ Unsupervised
- ☐ Reinforcement Learning

## Question 12

1.5 / 1.5 points

Which of the following is true? (More than one might apply)

- ✓ ☐ Machine Learning is a subset of Deep learning
- ✓ ☐ Ideally, the dataset should be split into three sections; training, validation and testing sets
- ✓ ☐ Predicting a continuous output value for a given set of input is known as a classification problem
- ✓ ☐ None of these
- ✓ ☐ Predicting a continuous output value for a given set of input is known as a regression problem

## Question 13

0.75 / 0.75 points

Inferencing is the process of making predictions on the training data using the trained model.

- ☐ True
- ✓ ☒ False

## Question 14

1.5 / 1.5 points

Consider k-fold cross validation applied to a training dataset with 256 samples with two different values of k given by  $2^{k_1}$  and  $32\log_8 k_2$ . Thus we are effectively doing different cross validation experiments, one is  $2^{k_1}$ -fold cross validation and other is  $32\log_8 k_2$ -fold cross validation. What is the value of  $k_1$  and  $k_2$  such that both the cross validation experiments are examples of leave-one-out cross validation ?

Value of  $k_1$ :

\_\_\_8\_\_\_ ✓(50 %) Value of  $k_2$  : \_\_\_16777216\_\_\_ ✓(50 %)

## Question 15

0.75 / 0.75 points

Which of the following is the correct relationship between bias, variance and mean squared error (MSE)?

- ☐  $MSE = Bias^2 + Variance^2 + Irreducible Error$
- ☐  $MSE = Bias + Variance + Irreducible Error$
- ☐  $MSE = Bias + Variance^2 + Irreducible Error$

✓ ☐  $MSE = Bias^2 + Variance + Irreducible Error$

## Question 16

0.5 / 0.5 points

Overfitting means that the model performs well on test/unseen data but does not generalize well to training data.

- ☐ True
- ✓ ☐ False

## Question 17

1.5 / 1.5 points

Which of the following characteristics typically apply to a complex model? (Select all that apply)

- ✓ ☐ Low variance
- ✓ ☐ High bias
- ✓ ☐ High variance
- ✓ ☐ Low bias

## Question 18

1.5 / 1.5 points

Here is an unordered list of different steps in a machine learning model lifecycle.

Model training, Monitoring, Model deployment, Retraining, Data preprocessing.

List these steps in correct order below, starting from the first step. (There should be no typos and write the step names exactly as listed above.)

- ✓ \_\_2\_\_ Model Training
- ✓ \_\_3\_\_ Model Deployment
- ✓ \_\_1\_\_ Data Preprocessing
- ✓ \_\_4\_\_ Monitoring
- ✓ \_\_5\_\_ Retraining

## Question 19

0.75 / 1.5 points

A linear regression model between sales of a product (Y) and advertising budget (X) reported the following statistics:  $R^2=0.6$  and TSS(Total Sum of Squares) = 0.9. Select all that is true about this model. (Select all that apply)

- ✗ ☐ The proportion of variability in Y that can be explained using X from the model is 0.54.
- ➡ ✓ ☐ The amount of variability in sales not explained by the model is 0.36.

- ✓ ☐ The amount of variability in the sales data that is explained by performing the regression is 0.36.
- ✓ ☐ The amount of variability inherent in the sales data before the regression is performed is 0.45.

**Question 20****0.5 / 0.5 points**

Underfitting means that the model is not complex enough to capture patterns in the training data well and therefore suffers from low performance on unseen data.

- ✓ ☐ True
- ☐ False

Done