Quiz 0 - Results

Attempt 1 of 1

Written Jan 22, 2024 5:33 PM - Jan 22, 2024 6:28 PM

Released Jan 24, 2024 2:00 PM

Attempt Score **58.2 / 70 - B**

Overall Grade (Highest Attempt) **58.2 / 70 - B**

Background Questionnaire

The first 4 survey questions are completion-based. All answer will receive full credit.

Question 1 **1 / 1 point**

Have you taken CSCI-UA 473 Fundamental of Machine Learning (previously named Introduction to Machine Learning) at NYU?

- ✔ ◯ Yes, I've taken it
- ◯ Not yet, I am taking it this semester
- ◯ No, I will take it in the future
- ◯ No, I don't plan to take it
- ◯ No, I've taken another equivalent course

Question 2 **1 / 1 point**

Rate your level of knowledge in Machine Learning in general: (5 is highest)

- ◯ 5
- ✖ ◯ 4
- ◯ 3
- ◯ 2

1

Question 3 **3 / 3 points**

Select the following Machine Learning topics which you are familiar with. Don't worry if you are unfamiliar with certain topics, we will adjust our schedule based on your answers:

- ✗ ☐ Linear Regression
- ✗ ☐ Regularization
- ✗ ☐ Logistic Regression
- ✗ ☐ Bias-variance tradeoff, Overfitting and Underfitting
- ✗ ☐ Cross-validation
- ✗ ☐ Hyperparameter Optimization
- ✗ ☐ Feature Engineering
- ✗ ☐ Model performance evaluation
- ✓ ☐ Perceptron
- ✓ ☐ Gradient Descent
- ✓ ☐ Training Neural Network in Tensorflow/Pytorch
- ✓ ☐ Convolutional Neural Networks
- ✓ ☐ Recurrent Neural Networks
- ✗ ☐ Reinforcement Learning

Question 4 **1 / 1 point**

Besides the topics listed in the syllabus and schedule posted on Brightspace, are there any other relevevnt topics that you are interested in and would like the course to cover more?
Feel free to enter "N/A" if you do not want to provide any response

N/A

The correct answer is not displayed for Written Response type questions.

Quiz 0

Question 5 **0 / 1 point**

You've just finished training a decision tree for spam classification, and it is getting abnormally bad performance on both your training and test sets. You know that

your implementation has no bugs, so what could be causing the problem?

- ➡ ◯ Your decision trees are too shallow
- ◯ You need to increase the learning rate
- ◯ You are overfitting
- ✖ ◯ None of the above

Question 6 **0 / 1 point**

Which of the following is true in Python?

- ◯ A) If data is an ndarray, the index must be the same length as data.
- ✖ ◯ B) Series is a one-dimensional labeled array capable of holding any data type.
- ➡ ◯ C) Both A and B
- ◯ D) None of the above

Question 7 **1 / 1 point**

Identify the type of learning in which labeled training data is used.

- ✔ ◯ Supervised learning
- ◯ Semi unsupervised learning
- ◯ Reinforcement learning
- ◯ Unsupervised learning

Question 8 **0 / 1 point**

In building a linear regression model for a particular data set, you observe the coefficient of one of the features having a relatively high negative value. This suggests that

- ✖ ◯ This feature has a strong effect on the model (should be retained)
- ◯ This feature does not have a strong effect on the model (should be ignored)
- ➡ ◯ It is not possible to comment on the importance of this feature without additional information

- ◯ None of the above

Question 9 **1 / 1 point**

The prior goal of the unsupervised learning model is to determine the _____ .

- ✔ ◯ Data patterns
- ◯ Classification task
- ◯ Discuss
- ◯ Regression task

Question 10 **1 / 1 point**

The error function most suited for gradient descent using logistic regression is

- ◯ The number of mistakes.
- ✔ ◯ The cross-entropy function.
- ◯ The squared error.
- ◯ The entropy function.

Question 11 **1 / 1 point**

Which of the following is NOT a preprocessing step in Scikit-learn?

- ✔ ◯ Regularization
- ◯ Encoding
- ◯ Scaling
- ◯ Imputation

Question 12 **0 / 1 point**

Which of the following options can be used to get global minima in K-means Algorithm?

- ✖ ◯ Try to run an algorithm for different centroid initialization
- ◯ Adjust the number of iterations
- ◯ Find out the optimal number of clusters
- ➡ ◯ All of the above

Question 13 **1 / 1 point**

What is the purpose of the GridSearchCV function in Scikit-learn?

- ✔ To tune the hyperparameters of a model
- To preprocess the data for modeling
- To evaluate the performance of a model
- To select the best features for modeling

Question 14 **1 / 1 point**

Logistic regression is used when you want to:

- Predict a continuous variable from dichotomous or continuous variables.
- ✔ Predict a dichotomous variable from continuous or dichotomous variables.
- Predict any categorical variable from several other categorical variables.
- Predict a continuous variable from dichotomous variables.

Question 15 **1 / 1 point**

Predicting whether it will rain or not tomorrow evening at a particular time is a type of _____ problem.

- ✔ Classification
- Regression
- Unsupervised learning
- All of the above

Question 16 **1 / 1 point**

Which evaluation metrics can be used to evaluate a model while modeling a continuous output variable?

- Accuracy
- ✔ Mean-Squared-Error
- AUC-ROC
- Log loss

Question 17 **0 / 1 point**

Adding a non-important feature to a linear regression model may result in.

- a) Increase in R-square
- b) Decrease in R-square
- c) Both (a) and (b)
- d) None of the above

Question 18 **1.2 / 2 points**

Which of the following is not supervised learning?

- [ ] Singular Value Decomposition (SVD)
- [x] Linear Regression
- [x] SVM
- [x] Principal Component Analysis (PCA)
- [x] Decision Tree

Question 19 **1 / 1 point**

Given a large dataset of medical records from patients suffering from heart disease, try to learn whether there might be different clusters of such patients for which we might tailor separate treatments. What kind of learning problem is this?

- a) Supervised learning
- b) Unsupervised learning
- c) Both (a) and (b)
- d) Neither (a) nor (b)

Question 20 **1 / 1 point**

The K-means algorithm:

- ✔ Minimizes the within-class variance for a given number of clusters
- Has the smallest value of the objective function when K = 1
- Converges to the global optimum if and only if the initial means are chosen as some of the samples themselves
- Requires the dimension of the feature space to be no bigger than the number of samples

Question 21 **1 / 1 point**

In Pandas, the key data structure is called?

- ✔ DataFrame
- Table
- Keyframe
- Tibble

Question 22 **0 / 1 point**

When performing regression or classification, which of the following is the correct way to preprocess the data?

- ➡ Normalize the data -> PCA -> training
- PCA -> normalize PCA output -> training
- ✖ Normalize the data -> PCA -> normalize PCA output -> training
- All of the above

Question 23 **1 / 1 point**

Consider that you are analyzing an extensive collection of fraudulent credit card transactions to discover if there are sub-types of these transactions. Which of the following learning methods best describes the given learning problem?

- Supervised Learning
- Semi-supervised learning
- Reinforcement Learning
- ✔ Unsupervised Learning

Question 24 **1 / 1 point**

Logistic regression assumes a:

- ✓ ◯ Linear relationship between continuous predictor variables and the logit of the outcome variable.
- ◯ Linear relationship between continuous predictor variables.
- ◯ Linear relationship between continuous predictor variables and the outcome variable.
- ◯ Linear relationship between observations.

Question 25 **1 / 1 point**

Which of the following methods prevent a model from overfitting to the training set?

- ◯ Dropout
- ◯ Early stopping
- ◯ Data augmentation
- ✓ ◯ All of these

Question 26 **1 / 2 points**

Which of the following methods can achieve zero training error on any linearly separable dataset?

- ✓ ☐ Logistic regression
- ✓ ☐ 15-nearest neighbors
- ➡ ✓ ☐ Perceptron
- ➡ ✗ ☐ Decision tree

Question 27 **1 / 1 point**

Which of the following is/are common uses of RNNs?

- ◯ Provide a caption for images
- ◯ Detect fraudulent credit card transaction
- ◯ Businesses help securities traders generate analytic reports
- ✓ ◯ All of the above

Question 28 **1 / 1 point**

Which, if any, of the following propositions is true about fully-connected neural networks (FCNN)?

- ◯ In an FCNN, there are connections between neurons of the same layer.
- ◯ An FCNN could only have linear activations.
- ◯ In an FCNN, the most common weight initialization scheme is zero initialization, because it leads to faster and more robust training.
- ✔ ◯ None of the above

Question 29 **1 / 1 point**

In which cases will K-means clustering fail to give good results?

- ◯ Data points with outliers
- ◯ Data points with different densities
- ◯ Data points with nonconvex shapes
- ✔ ◯ All of the above

Question 30 **0 / 1 point**

In linear regression, we try to _____ the _____ of the model to identify the line of best fit.

- ◯ Maximize; least square errors
- ◯ Maximize; residuals
- ✖ ◯ Minimize; residuals
- ➡ ◯ Minimize; least square errors
- ◯ None of the above

Question 31 **0 / 1 point**

What is the disadvantage of decision trees?

- ◯ Factor analysis
- ✖ ◯ Decision trees are robust to outliers
- ➡ ◯ Decision trees are prone to be overfit
- ◯ All of the above

Question 32 **1 / 1 point**

Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging?

- ◯ Regression
- ◯ Decision Tree
- ◯ Classification
- ✔◯ Random Forest

Question 33 **1 / 1 point**

Lasso can be interpreted as least-squares linear regression where

- ✔◯ Weights are regularized with the L1 norm
- ◯ The weights have a Gaussian prior
- ◯ Weights are regularized with the L2 norm
- ◯ The solution algorithm is simpler

Question 34 **1 / 1 point**

Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of n classes (p1, p2.., pk) such that the sum of p over all n equals to 1?

- ◯ Sigmoid
- ◯ Tanh
- ◯ ReLu
- ✔◯ Softmax

Question 35 **1 / 1 point**

What does a gradient descent algorithm do?

- ◯ A) Tries to find the parameters of a model that minimizes the cost function
- ◯ B) Update the weights between layers
- ✔◯ C) Both A and B

D) None

Question 36 **0.5 / 1.5 points**

Which of the following statement(s) is true for Gradient Descent (GD) and Stochastic Gradient Descent (SGD)?

- ✓ In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.
- ✓ In SGD, you must run through all the samples in your training set for a single parameter update in each iteration.
- ✗ In GD, you either use the entire data points or a subset of training data to update a parameter in each iteration.

Question 37 **1 / 1 point**

Real-time decisions, Game AI, Learning Tasks, Skill acquisition, and Robot Navigation are applications of_____

- ✓ Reinforcement Learning
- Supervised Learning: Classification
- Supervised Learning: Regression
- Unsupervised Learning
- None of the above

Question 38 **1 / 1 point**

Which of the following is NOT a classification metric used in Scikit-learn?

- F1-score
- Recall
- ✓ R-squared
- Precision

Question 39 **1 / 1 point**

Which of the following is the correct order for the Convolutional Neural Network operation?

- ✓ Convolution -> max pooling -> flattening -> full connection

- ◯ Max pooling -> convolution -> flattening -> full connection
- ◯ Flattening -> max pooling -> convolution -> full connection
- ◯ None of them

Question 40 **1 / 1 point**

What would you do in PCA to get the same projection as SVD?

- ✔ ◯ Transform data to zero mean
- ◯ Transform data to zero median
- ◯ Not possible
- ◯ None of these

Question 41 **1 / 1 point**

Netflix maintains a dataset of its members' information that includes their locations and watch history. Now, based on these records, the company wants to divide the members into seven groups so that it can push different ads to them. Which of the following methods should Netflix use to get the groups?

- ◯ k-nn
- ◯ Convolutional neural network
- ✔ ◯ k-means
- ◯ Linear regression

Question 42 **1 / 1 point**

Suppose we have a regularized linear regression model. What is the effect of increasing λ on bias and variance?

- ◯ Decreases bias, increases variance
- ◯ Decreases bias, decreases variance
- ◯ Increases bias, increases variance
- ✔ ◯ Increases bias, decreases variance

Question 43 **1 / 1 point**

Which of the following is a method for evaluating the performance of a classification model in Scikit-learn?

- [x] Accuracy
- [ ] Root mean squared error (RMSE)
- [ ] R-squared
- [ ] Mean squared error (MSE)

Question 44 **3.5 / 3.5 points**

Match the purpose of the following methods in Scikit-learn:

✓ __6__	To compute the probability estimates of the class labels
✓ __1__	To train a model using a given dataset
✓ __2__	To make predictions using a trained model
✓ __7__	To transform the predicted values back to their original scale
✓ __3__	To evaluate the performance of a model
✓ __5__	To preprocess the data for modeling
✓ __4__	To visualize the data using a plot

1. fit()
2. predict()
3. score()
4. plot()
5. transform()
6. predict_proba()
7. inverse_transform()

Question 45 **1 / 1 point**

Which of the following is a disadvantage of the k-fold cross-validation method?

- [ ] This usually does not take longer time to compute
- [x] The training algorithm has to rerun from scratch k times
- [ ] The variance of the resulting estimate is reduced as k is increased.
- [ ] Reduced bias

Question 46 **2 / 2 points**

Regarding bias and variance, which of the following statements are true? (Here 'high' and 'low' are relative to the ideal model.)

- [x] Models that underfit have a low variance

- [x] Models that overfit have a low bias.
- [x] Models that overfit have a high bias.
- [x] Models that underfit have a high variance.

Question 47 **1 / 1 point**

Which of the following neural network models has a shared weight structure?

- A) Recurrent Neural Network
- B) Convolution Neural Network
- [x] C) Both A and B
- D) None

Question 48 **1 / 1 point**

Which of the following Numpy operations are correct?

- Mathematical and logical operations on arrays.
- Fourier transforms and routines for shape manipulation.
- Operations related to linear algebra.
- [x] All of the above

Question 49 **1 / 1 point**

Consider a point that is correctly classified and distant from the decision boundary. Which of the following methods will be unaffected by this point?

- [x] SVM
- Logistic regression
- Nearest neighbor
- Linear regression

Question 50 **1 / 1 point**

You trained a binary classifier model which gives very high accuracy on the training data, but much lower accuracy on validation data. Which of the following may NOT be true?

- ◯ The training and testing examples are sampled from different distributions.
- ✔◯ This is an instance of underfitting.
- ◯ This is an instance of overfitting.
- ◯ The training was not well regularized.

Question 51 **1 / 1 point**

Which of the following is a method for selecting the best hyperparameters for a model in Scikit-learn?

- ◯ Grid search
- ◯ Random search
- ◯ Bayesian optimization
- ✔◯ All of the above

Question 52 **1 / 1 point**

Which of the following are correct features of DataFrame?

- ◯ Potentially columns are of different types
- ◯ Can Perform Arithmetic operations on rows and columns
- ◯ Labeled axes (rows and columns)
- ✔◯ All of the above

Question 53 **1 / 1 point**

_____ refers to a model that can neither model the training data nor generalize to new data.

- ◯ good fitting
- ◯ overfitting
- ✔◯ underfitting
- ◯ All of the above

Question 54 **1 / 1 point**

Given a Neural Net with N input nodes, no hidden layers, one output node, with Entropy Loss and Sigmoid Activation Functions, which of the following algorithms (with the proper hyper-parameters and initialization) can be used to find the global optimum?

- ◯ Stochastic Gradient Descent
- ◯ Mini-Batch Gradient Descent
- ◯ Batch Gradient Descent
- ✔ ◯ All of the above

Question 55 **1 / 1 point**

Which of the following are the most widely used metrics and tools to assess a classification model?

- ◯ Confusion matrix
- ◯ Cost-sensitive accuracy
- ◯ Area under the ROC curve
- ✔ ◯ All of the above

Question 56 **1 / 1 point**

Which of the following can be data in Pandas?

- ◯ a dictionary
- ◯ an ndarray
- ◯ a scalar value
- ✔ ◯ All of the above

Question 57 **1 / 1 point**

The most important object defined in NumPy is an N-dimensional array type called?

- ◯ nd_array
- ✔ ◯ ndarray
- ◯ narray
- ◯ darray

Question 58 **1 / 1 point**

In many real-world classification problems, the datasets are often nonlinearly separable. When using a neural network to solve such problems, we need to add nonlinearity to the network. Which of the following will add nonlinearity to the network?

- ✔ ◯ using a rectified linear function as an activation function
- ◯ adding a hidden layer
- ◯ removing a hidden layer
- ◯ using a linear function as the activation function

Question 59 **1 / 2 points**

Which of the following can ONLY be used when training data are linearly separable?

- ✔ ☐ Logistic Regression
- ➡ ✔ ☐ Hard-margin SVM
- ✔ ☐ Soft margin SVM
- ➡ ✖ ☐ Perceptron

Question 60 **1 / 1 point**

For an image classification task, which of the following deep learning architectures is best suited?

- ◯ Recurrent Neural Network
- ◯ Multi-Layer Perceptron
- ✔ ◯ Convolution Neural Network
- ◯ All of the above

Question 61 **1 / 1 point**

In Reinforcement Learning, ϵ-greedy algorithm describes the tradeoff between

- ✔ ◯ exploration and exploitation
- ◯ agent and environment
- ◯ action and reward
- ◯ the agent's action and the environment's state

Question 62 **0 / 0.0001 points**

Please select all the time slots you are eligible for attending OH. Note that for project discussion in the future, you will have to attend OH for advisement.

- ✗ ☐ Mondays 12 - 2 PM
- ✗ ☐ Mondays 2 - 4 PM
- ✓ ☐ Tuesdays 10 AM - 12 PM
- ✓ ☐ Tuesdays 12 - 2 PM
- ✓ ☐ Tuesdays 2 - 4 PM
- ✗ ☐ Wednesdays 10 AM - 12 PM
- ✗ ☐ Wednesdays 12 - 2 PM
- ✗ ☐ Wednesdays 2 - 4 PM
- ✓ ☐ Thursdays 12 - 2 PM
- ✓ ☐ Thursdays 2 - 4 PM
- ✓ ☐ Fridays 10 AM - 12 PM

Done