Review Test Submission: Week 1 Quiz

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Week 1 Quiz
Started	9/2/18 9:48 AM
Submitted	9/2/18 10:02 AM
Due Date	9/2/18 11:59 PM
Status	Completed
Attempt Score	24 out of 25 points
Time Elapsed	14 minutes out of 15 minutes
Instructions	For each question select the answer that is most closely correct based on your readings and discussions in the class. There are 25 questions. You will have 15 minutes to complete the quiz.

You will have one opportunity to take the quiz and once you hit SUBMIT you will not have an opportunity to change any of your answers. Please check you work before submitting.

This quiz is autograded. There is no partial credit. Each question is worth the same amount of your grade.

Make sure you have a stable internet connection before starting. Hardwired is preferable to wi-fi. Also, you should close all other applications before beginning.

If you encounter any errors or situations that prevent you from completing the quiz please try to document it by taking a screencap or printscrn. |
| Results Displayed | All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions |

Question 1

1 out of 1 points

In machine learning, the term supervised refers to a set of samples where the desired output signals (labels) are already known.

Selected Answer: ✅ True

Answers: ✅ True

                  False

Question 2

1 out of 1 points

The 'curse of dimensionality' refers to :

Selected Answer: ✅ the excessive number of features associated with a particular instance

Answers: high volumes of observational data

the overreliance on multivariate linear models

the tendency of the ratio of features to observations to remain constant

✅ the excessive number of features associated with a particular instance

Question 3

A type of supervised learning that is for the prediction of continuous numerical outcomes values is more commonly called called regression.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 4

_____ refers to the phenomenon present in every machine learning method, namely that when the number of parameters of a model grows, the data needed to learn them grows exponentially.

Selected Answer: ✅ The curse of dimensionality

Answers: The Idem rule

✅ The curse of dimensionality

Hyperparameter tuning

Stochastic Gradient Descent

Question 5

_____ can be used to compress the data onto a smaller dimensional subspace while retaining most of the relevant information.

Selected Answer: ✅ dimensionality reduction

Answers: data compression

✅ dimensionality reduction

data normalization

bias/variance trade-off

data standardization

Question 6

According to Garreta and Moncecchi, which of the following are important concepts related to machine learning:

Selected Answer: ✅ all of the above are important concepts

Answers: overfitting/underfitting

Occam's razor

bias/variance trade-off

feature standardization and normalization

feature engineering

✅ all of the above are important concepts

Question 7

1 out of 1 points

Clustering is also sometimes called unsupervised classification.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 8

1 out of 1 points

The two general subcategories of supervised learning are _____ and _____.

Selected Answer: ✅ classification and regression

Answers: classification and clustering

clustering and dimensionality reduction

✅ classification and regression

classification and reinforcement

regression and clustering

Question 9

0 out of 1 points

In regression analysis, we are given a number of continuous response variables (outcomes or targets) and a predictor (explanatory) variable, and we try to find a relationship between those variables that allows us to make a prediction.

Selected Answer: ❌ True

Answers: True

✅ False

Question 10

1 out of 1 points

The machine learning roadmap describes four stages in predictive modeling. Select the four from the following list:

Selected Answers: ✅ Prediction

	✅ Learning
	✅ Preprocessing
	✅ Evaluation
Answers:	✅ Prediction
	Reweighting
	Heuristic development
	✅ Learning
	Back Propagation
	Analysis
	Optimization
	✅ Preprocessing
	✅ Evaluation

Question 11

1 out of 1 points

Some examples of machine learning in our everyday lives include:

Selected Answer: ✅ all of the above

Answers: email spam filters

text and voice recognition

web search engines

chess playing programs

✅ all of the above

Question 12

1 out of 1 points

A supervised learning task with discrete class labels, such as in the previous email spam filtering example, is also called a _____ task.

Selected Answer: ✅ classification

Answers: association

✅ classification

clustering

regression

reinforcement

Question 13

1 out of 1 points

If a dataset contains class labels, it can only be used for regression analysis.

Selected Answer: ✅ False

Answers: True

✅ False

Question 14

1 out of 1 points

Attributes, measurements and features all refer to the same thing and may be used interchangeably.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 15

1 out of 1 points

In machine learning, the term supervised refers to a process in the design stage in which a model is trained directly under the supervision of a human expert.

Selected Answer: ✅ False

Answers: True

✅ False

Question 16

1 out of 1 points

Linear regression is so called because it makes use of strictly linear decision boundaries between multiclass variables.

Selected Answer: ✅ False

Answers: True

✅ False

Question 17

1 out of 1 points

Each of the following are a consideration in the use of dimensionality reduction except:

Selected Answer: ✅ data normalization/standardization

Answers: ✅ data normalization/standardization

limited storage space

algorithmic computational performance

removal of noise in the data or excessive correlations between features

data visualization techniques

Question 18

1 out of 1 points

In the second half of the twentieth century, machine learning evolved as a subfield of Artificial Intelligence (AI) that involved self-learning algorithms that derived knowledge from data in order to make
predictions.

Selected Answer: ✅ True

Answers: ● True
                  False

Question 19

1 out of 1 points

Which of the following is not a typical example of a multiclass classification task:

Selected Answer: ● email spam filter

Answers: ● email spam filter

                    handwritten character recognition

                    Iris flower variety identification

                    MNIST handwritten digit recognition

Question 20

1 out of 1 points

A machine learning task in which an algorithm is trained on a corpus of labeled emails, emails that are labelled as spam or not-spam, to predict whether a new email belongs to either of the two categories is an example of:

Selected Answer: ● binary classification

Answers: multiclass classification

                    ● binary classification

                    unsupervised classification

                    clustering

                    multivariate regression

Question 21

1 out of 1 points

The main goal in unsupervised learning is to learn a model from labeled training data that allows us to make predictions about unseen or future data.

Selected Answer: ● False
Answers: True
                  ● False

Question 22

1 out of 1 points

In regression analysis, we are given a number of predictor (explanatory) variables and a continuous response variable (outcome or target), and we try to find a relationship between those variables that allows us to predict an outcome.

Selected Answer: ● True
Answers: ● True
                  False

Question 23

1 out of 1 points

After successfully installing Anaconda, we can install new Python packages using which of the following commands?

Selected Answer: ✅ conda install SomePackage

Answers: conda update SomePackage

pip install SomePackage

pip install SomePackage --upgrade

✅ conda install SomePackage

Question 24

1 out of 1 points

In a dataset, an instance may have many features, but a feature can have only one instance.

Selected Answer: ✅ False
Answers: True
✅ False

Question 25

1 out of 1 points

The three types of machine learning discussed in class include all of the following except:

Selected Answer: ✅ adaptive

Answers: supervised

unsupervised

reinforcement

✅ adaptive

Sunday, October 21, 2018 12:54:16 PM CDT

← **OK**

🏠   Course Modules   Module 2 - Classification   **Review Test Submission: Module 2 Quiz**

Review Test Submission: Module 2 Quiz

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Module 2 Quiz
Started	9/7/18 7:25 PM
Submitted	9/7/18 7:39 PM
Due Date	9/9/18 11:59 PM
Status	Completed
Attempt Score	25 out of 25 points
Time Elapsed	13 minutes out of 20 minutes
Instructions	

For each question select the answer that is most closely correct based on your readings and discussions in the class. There are 20 questions. You will have 20 minutes to complete the quiz.

You will have one opportunity to take the quiz and once you hit SUBMIT you will not have an opportunity to change any of your answers. Please check you work before submitting.

This quiz is autograded. There is no partial credit. Each question is worth the same amount of your grade.

Make sure you have a stable internet connection before starting. Hardwired is preferable to wi-fi. Also, you should close all other applications before beginning.

If you encounter any errors or situations that prevent you from completing the quiz please try to document it by taking a screencap or printscrn.

Results Displayed	All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions

Question 1

1.25 out of 1.25 points

Which of the following does NOT describe a k-nearest neighbor (KNN) classifier?

Selected Answer:	✅ eager learner
Answers:	non-parametric
	✅ eager learner
	lazy learner
	instance based classifier

Question 2

1.25 out of 1.25 points

When the value for K is set to be a large number (relative to the size of the training set), the training time for a KNN classifier:

Selected Answer: ✅ will remain at zero.

Answers: will increase linearly.

will increase exponentially.

will increase proportionally to the size of the training set.

✅ will remain at zero.

Question 3

1.25 out of 1.25 points

According to the Raschka textbook, which of the following is not an example of a paramentric machine learning model:

Selected Answer: ✅ random forest

Answers: ✅ random forest

perceptron

logistic regression

linear SVM

Question 4

1.25 out of 1.25 points

According to the Raschka textbook, which of the following is not an example of a non-paramentric machine learning model:

Selected Answer: ✅ logistic regression

Answers: decision trees

random forests

KNN

✅ logistic regression

Question 5

1.25 out of 1.25 points

Decision trees work equally well with features that are categorical (like "hot", "mild" or "cold") or continuous numeric (like temperature).

Selected Answer: ✅ True

Answers: ✅ True

False

Question 6

1.25 out of 1.25 points

Decision trees require that all variables are scaled to the same range, a process called normalization or standardization.

Selected Answer: ✅ False

Answers: True

✅ False

Question 7

1.25 out of 1.25 points

A decision tree may continue splitting the dataset at each node until all the instances at each leaf node are pure.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 8

1.25 out of 1.25 points

A decision tree may have many internal nodes and many root nodes, but only one leaf node.

Selected Answer: ✅ False

Answers: True

✅ False

Question 9

1.25 out of 1.25 points

In order to determine the next split, a decision tree algorithm will try to minimize Information Gain.

Selected Answer: ✅ False

Answers: True

✅ False

Question 10

1.25 out of 1.25 points

Theoretically, a decision tree may be either binary or non-binary, but in practice most are non-binary for ease of computation.

Selected Answer: ✅ False

Answers: True

✅ False

Question 11

1.25 out of 1.25 points

The following are all distance metrics used in nearest neighbor calculations except:

Selected Answer: ✅ Turing

Answers: ✅ Turing

Euclidean

Manhattan

Minkowski

Question 12

1.25 out of 1.25 points

According to the text, the Minkowski distance measure is a generalization of the _____ and _____ distance.

Selected Answer: ✅ Euclidean and Manhattan

Answers: Euclidean and Stammering

✅ Euclidean and Manhattan

Manhattan and Eulerean

Maharattan and Minkowski

Question 13

1.25 out of 1.25 points

If we are using a Euclidean distance measure, it is important to _____ the data so that each feature contributes equally to the distance.

Selected Answer: ✅ standardize

Answers: Winsorize

truncate

✅ standardize

regularize

Question 14

1.25 out of 1.25 points

Which of the following is not a standard measure of impurity used in constructing decision trees?

Selected Answer: ✅ Pruning ratio

Answers: Entropy

✅ Pruning ratio

classification error

Gini impurity

Question 15

1.25 out of 1.25 points

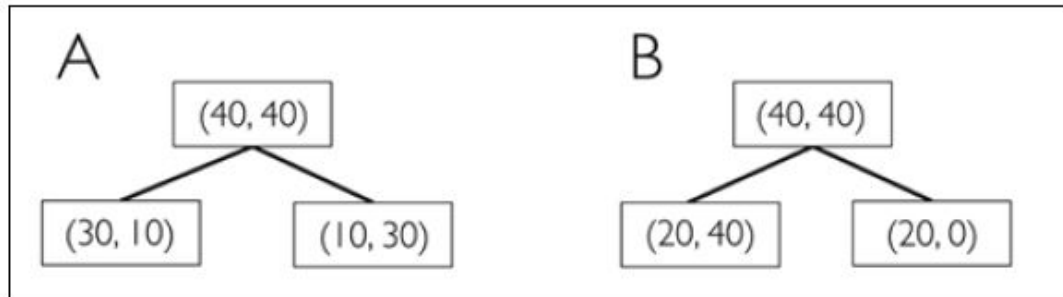Decision trees can build complex decision boundaries by dividing the feature space into rectangles.

Selected Answer: ✓ True

Answers: ✓ True

False

Question 16

1.25 out of 1.25 points

Use the figure on Page 91 of the Raschka text to answer the next five questions.

Starting with a dataset at the parent node, Dp which consists of 40 samples from Class 1 and 40 samples from Class 2.

Look at the two possible splitting scenarios (A and B) shown in the figure.

What is the classification error at the parent node?

Selected Answer: ✓ 0.5

Answers: 0.25

✓ 0.5

0.75

1.0

Question 17

1.25 out of 1.25 points

What is the Information Gain of scenario A using classification error?

Selected Answer: ✓ 0.25

Answers: ✓ 0.25

0.5

0.75

1.0

Question 18

1.25 out of 1.25 points

What is the Gini impurity of the right branch in Scenario B?

Selected Answer: ✅ 0.0

Answers: ✅ 0.0

0.25

0.375

0.5

Question 19

1.25 out of 1.25 points

What is the Information Gain for Scenario A using Gini impurity?

Selected Answer: ✅ 0.125

Answers: 0.0

✅ 0.125

0.25

0.5

Question 20

1.25 out of 1.25 points

Using Entropy criterion which scenario has the highest Information Gain?

Selected Answer: ✅ Scenario B

Answers: Scenario A

✅ Scenario B

They are both equal

They are both equal to 1

Sunday, October 21, 2018 12:55:33 PM CDT

← OK

Review Test Submission: Module 3 Quiz

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Module 3 Quiz
Started	9/16/18 3:29 PM
Submitted	9/16/18 3:44 PM
Due Date	9/16/18 11:59 PM
Status	Completed
Attempt Score	17.5 out of 25 points
Time Elapsed	15 minutes out of 15 minutes
Instructions	This quiz may contain questions of the following types: True/false, multiple choice, fill in the blank, short answer or matching. It is open book and open notes, but you may not consult with anyone while taking the test. It should be your own work.

It consists of 20 questions worth an equal amount each. It is timed to 15 minutes. It must be completed in one sitting. It will auto-submit at the end of the timed interval automatically. |
| Results Displayed | All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions |

Question 1
1.25 out of 1.25 points

Each row in a dataset represents an individual case (also called an instance, example, or observation).

Selected Answer:	✅	True
Answers:	✅	True
		False

Question 2
1.25 out of 1.25 points

Attributes (the variables being used to make predictions) are also known as the following (Select the answer that is most correct). :

Selected Answer:	✅	all of the above are correct
Answers:		predictors
		features
		independent variables

inputs

    ✅ all of the above are correct

Question 3

1.25 out of 1.25 points

Class labels are also known as the following (Slect the answer that is most correct):

Selected Answer: ✅ All of the above are correct

Answers: Outcomes

    Targets

    Dependent variables

    Responses

    ✅ All of the above are correct

Question 4

1.25 out of 1.25 points

If a dataset includes a categorical or factor class label, which of the following would apply:

Selected Answer: ✅ all of the above could apply

Answers: supervised learning

    classification task

    k-nearest neighbors

    decision tree

    ✅ all of the above could apply

Question 5

1.25 out of 1.25 points

When conducting EDA on a new dataset which of the following is not something that you should check?

Selected Answer: ✅ You should check all of these.

Answers: Number of rows and columns

    Number of categorical variables and number of unique values for each

    Missing values

    Summary statistics for attributes and labels

    ✅ You should check all of these.

Question 6

1.25 out of 1.25 points

Categorical variables can be two-valued, like Male Female, or multivalued, like states (AL, AK, AR . . . WY).

Selected Answer: ✅ True
Answers: ✅ True
False

Question 7

1.25 out of 1.25 points

One should never convert a continuous numeric data label (for example, $120, $200, $450, $900) into a categorical value (like if >$200 - True or False).

Selected Answer: ✅ False
Answers: True
✅ False

Question 8

1.25 out of 1.25 points

There are several ways of dealing with missing values in your dataset. Which of these is the least favorable?

Selected Answer: ✅ Delete all instances with missing values in any attribute

Answers: ✅ Delete all instances with missing values in any attribute

Use mean imputation to replace all missing values

Use median imputation to replace all missing values

Use most_frequest imputation to replace all missing values

Question 9

1.25 out of 1.25 points

In a Q-Q plot, if the data being analyzed comes from a Gaussian distribution, the point being plotted will lie on a straight line.

Selected Answer: ✅ True
Answers: ✅ True
False

Question 10

1.25 out of 1.25 points

A scatter plot is a quick and effective way to describe visually the relationship between two numeric attributes.

Selected Answer: ✅ True
Answers: ✅ True
False

Question 11

1.25 out of 1.25 points

The degree of correlation between two attributes (or an attribute and a target) can be quantified using Pearson's correlation coefficient.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 12

1.25 out of 1.25 points

Pearson's correlation coefficient takes on a value between 0 and 1, where 0 indicates no correlation between the attributes and 1 indicates perfect correlation.

Selected Answer: ✅ False

Answers: True

✅ False

Question 13

1.25 out of 1.25 points

One way to check correlations with a large number of attributes is to calculate the Pearson's correlation coefficient for all pairs of attributes, arrange those correlations into a matrix, and then plot them in a _____.

Selected Answer: ✅ heat map

Answers: bee swarm plot

✅ heat map

target-attribute crossplot

box plot

Question 14

1.25 out of 1.25 points

How would you complete the following Python code:

#calculate correlations between real-valued attributes
corMat = DataFrame(rocksVMines._____())

Selected Answer: ✅ corr

Answers: plot

✅ corr

show

describe

Question 15

0 out of 1.25 points

How would you complete the following Python code:
#read rocks versus mines data into pandas data frame
target_url = ("https://archive.ics.uci.edu/ml/machine-learning-""databases/undocumented/connectionist-bench/sonar/sonar.all-data")

rocksVMines = pd._____(target_url,header=None, prefix="V")

Selected Answer: ❌ [None Given]

Answers: load_file

read_input

get_data

✅ read_csv

Response Feedback: read_csv is correct in this example.

Question 16

0 out of 1.25 points

ECDF is a commonly used statistical term which stands for Experimental Continuous Density Function.

Selected Answer: ❌ [None Given]

Answers: True

✅ False

Response Feedback: empirical cumulative distribution function

Question 17

0 out of 1.25 points

When constructing a histogram it is necessary to be aware of binning bias to avoid misleading comparisons by changing the number or width of bins.

Selected Answer: ❌ [None Given]

Answers: ✅ True

False

Response
Feedback: binning bias is a problem associated with histogram construction, mentioned in the Datacamp chapter.

Question 18

0 out of 1.25 points

Which of the following quotes is not directly from John W. Tukey?

Selected
Answer: ❌ [None Given]

Answers: Exploratory data analysis is detective work.

There is no excuse for failing to plot and look.

The greatest value of a picture is that it forces us to notice what we never expected to see.

It is important to understand what you can do before you learn how to measure how well you seem to have done it.

✅
Often times EDA is too time consuming, so it is better to jump right in and do your hypothesis tests.

Response Feedback: "Often times EDA is too time consuming, so it is better to jump right in and do your hypothesis tests." This is the opposite of what John Tukey would say!

Question 19

0 out of 1.25 points

Which of the following is not true of graphical EDA?

Selected Answer: ❌ [None Given]

Answers: It often involves converting tabular data into graphical form.

If done well, graphical representations can allow for more rapid interpretation of data.

A nice looking plot is always the end goal of a statistical analysis.

✅

There is no excuse for neglecting to do graphical EDA.

Response Feedback: "A nice looking plot is always the end goal of a statistical analysis." It is the starting point, but not the end goal. (From the Datacamp chapter.)

Question 20

0 out of 1.25 points

Select all of the following that are true statements about John W. Tukey:

Selected Answers: ✅ invented the statistical technique known as the Box Plot
                    ❌ won the Nobel Prize for statistics while at Princeton University
                    ✅ coined the term 'bit' for binary digit
                    ✅ published the textbook "Exploratory Data Analysis" in 1977
                    ✅ worked with both John von Neumann and Claude Shannon

Answers: ✅ invented the statistical technique known as the Box Plot
           won the Nobel Prize for statistics while at Princeton University
           ✅ coined the term 'bit' for binary digit
           ✅ published the textbook "Exploratory Data Analysis" in 1977
           ✅ worked with both John von Neumann and Claude Shannon
           penned an influential 'think piece' in 1962 called "The End of Data Analysis"

Response Feedback: He never won the Nobel Prize and the title of the article he wrote was "The Future of Data Analysis."

Sunday, October 21, 2018 12:56:28 PM CDT

🏠 Course Modules   Module 4 - Linear and Penalized Linear Regression

Review Test Submission: Module 4 Quiz

Review Test Submission: Module 4 Quiz

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Module 4 Quiz
Started	9/21/18 7:17 PM
Submitted	9/21/18 7:29 PM
Due Date	9/23/18 11:59 PM
Status	Completed
Attempt Score	24 out of 25 points
Time Elapsed	11 minutes out of 20 minutes
Instructions	This quiz may contain questions of the following types: True/false, multiple choice, fill in the blank, short answer or

matching. It is open book and open notes, but you may not consult with anyone while taking the test. It should be your
own work.

It consists of 25 questions worth an equal amount each. It is timed to 20 minutes. It must be completed in one sitting.
It will auto-submit at the end of the timed interval automatically. |
| Results Displayed | All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions |

Question 1

1 out of 1 points

Which of the following is a false statement?

Selected Answer:	✅ Ridge is useful as a supervised feature selection technique.
Answers:	Using Lasso, depending on the regularization strength, certain weights can become zero.
	Ridge regression is an L2 penalized model where we simply add the squared sum of
the weights to our least-squares cost function.
	✅ Ridge is useful as a supervised feature selection technique.
	By increasing the value of hyperparameter lambda, we increase the regularization
strength and shrink the weights of our model.

Question 2

1 out of 1 points

Lasso regression is an L1 penalized model where we simply add the absolute value of the weights to our least-squares cost function.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 3

1 out of 1 points

Overfitting is a common problem in machine learning, where a model performs well on training data but does not generalize well to unseen data (test data).

Selected Answer: ✅ True

Answers: ✅ True

False

Question 4

1 out of 1 points

_____ can be used for feature selection.

Selected Answer: ✅ LASSO

Answers: ✅ LASSO

univariate regression

Ridge

OLS

L2 regularization

Question 5

1 out of 1 points

By decreasing the value of hyperparameter lambda, we increase the regularization strength and shrink the weights of our model.

Selected Answer: ✅ False

Answers: True

✅ False

Question 6

1 out of 1 points

If we set the l1_ratio to 0, the ElasticNet regressor would be equal to LASSO regression.

Selected Answer: ✅ False

Question 7

1 out of 1 points

If a model suffers from overfitting, we also say that the model has a high _____, which can be caused by having too many parameters that lead to a model that is too complex given the underlying data.
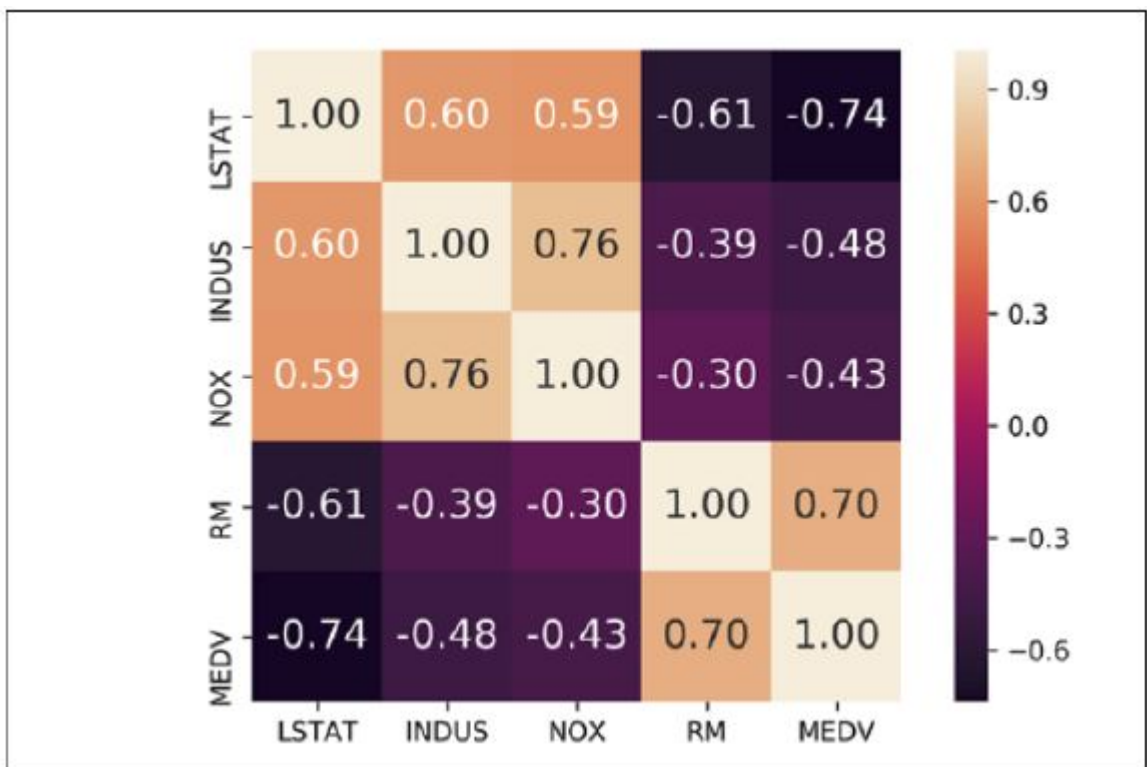
Selected Answer: ✅ variance

Answers: precision

recall

bias

✅ variance

Question 8

1 out of 1 points

Using this heatmap of the features from the Housing dataset, what is the correlation between Median Value and Number of Rooms?

Selected Answer: ✅ 0.70

Answers: -0.74

1.0

-0.61

✅ 0.70

Question 9

1 out of 1 points

A model can also suffer from _____ (high bias), which means that our model is not complex enough to capture the pattern in the training data well and therefore also suffers from low performance on unseen data.

Selected Answer: ✅ underfitting

Answers: ✅ underfitting

convergence

overfitting

sparcity

Question 10

1 out of 1 points

Elastic Net has an L1 penalty to generate sparsity and an L2 penalty to overcome some of the limitations of LASSO.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 11

1 out of 1 points

Using Ridge, depending on the regularization strength, certain weights can become zero.

Selected Answer: ✅ False

Answers: True

✅ False

Question 12

1 out of 1 points

When a model's performance is calculated on the data used to train the model this is called an "in sample" test. When it is calculated on unseen test data, it is called "out of sample" testing.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 13

1 out of 1 points

Regularization introduces additional information (bias) to penalize extreme parameter (weight) values.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 14

1 out of 1 points

Using Lasso can lead to sparse models.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 15

1 out of 1 points

Regularization is a very useful method to: (Select the one that does NOT apply)

Selected Answer: ✅ eliminate scarcity

Answers: prevent overfitting

filter out noise from data

handle collinearity (high correlation among features)
filter out noise from data, and
eventually prevent overfitting.

✅ eliminate scarcity

Question 16

1 out of 1 points

Linear regression can be understood as finding the best-fitting straight line through the sample points.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 17

1 out of 1 points

The target or response variable of the Housing dataset, used in the examples this week is:

Selected Answer: ✅ MEDV

Answers: AGE

✅ MEDV

LSTAT

RM

Question 18

1 out of 1 points

Lasso regression is an L2 penalized model where we simply add the squared sum of the weights to our least-squares cost function.

Selected Answer: ✅ False
Answers: True
       ✅ False

Question 19

1 out of 1 points

The most common form of regularization is L2 regularization (sometimes also called L2 shrinkage or weight decay).

Selected Answer: ✅ True
Answers: ✅ True
       False

Question 20

1 out of 1 points

Regression analysis aims to predict outputs on a continuous scale rather than categorical class labels.

Selected Answer: ✅ True
Answers: ✅ True
       False

Question 21

1 out of 1 points

The residual (or residual error) is equal to the vertical offset; $|y_{pred} - y_i|$.

Selected Answer: ✅ True
Answers: ✅ True
       False

Question 22

1 out of 1 points

Sparsity can be useful in practice if we have a high-dimensional dataset with many features that are irrelevant, especially cases where we have more irrelevant dimensions than samples.

Selected Answer: ✅ True
Answers: ✅ True
       False

Question 23

0 out of 1 points

For regularization to work properly, it is not necessary to ensure that all our features are on comparable scales.

Selected Answer: ❌ True

Answers: True

✅ False

Response
Feedback:

It is necessary for features to be on the same scales for regularization to work.

Question 24

1 out of 1 points

If we set the l1_ratio to 1.0, the ElasticNet regressor would be equal to LASSO regression.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 25

1 out of 1 points

The goal of multivariate linear regression is to model the relationship between a single feature (explanatory variable x) and a continuous valued response (target variable y).

Selected Answer: ✅ False

Answers: True

✅ False

Sunday, October 21, 2018 12:57:02 PM CDT

← **OK**

Review Test Submission: Week 5 QUIZ - Dimensionality Reduction and PCA

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Week 5 QUIZ - Dimensionality Reduction and PCA
Started	9/28/18 8:20 PM
Submitted	9/28/18 8:28 PM
Due Date	9/30/18 11:59 PM
Status	Completed
Attempt Score	25 out of 25 points
Time Elapsed	7 minutes out of 20 minutes
Instructions	This quiz may contain questions of the following types: True/false, multiple choice, fill in the blank, short answer or matching. It is open book and open notes, but you may not consult with anyone while taking the test. It should be your own work. It consists of 20 questions worth an equal amount each. (It looks like 14, but question #11 counts as 7 questions!) It is timed to 20 minutes. It must be completed in one sitting. It will auto-submit at the end of the timed interval automatically.
Results Displayed	All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions

Question 1

1.25 out of 1.25 points

_____ is an example of a feature selection technique:

Selected Answer: ✅ LASSO

Answers: PCA

SVM

LDA

✅ LASSO

Question 2

1.25 out of 1.25 points

_____ is the practice of summarizing the information content of a dataset by transforming it onto a new feature subspace of lower dimensionality than the original one.

Selected Answer: ✓ Feature extraction

Answers: Feature manipulation

       Feature expansion

       ✓ Feature extraction

       Feature inflation

Question 3

1.25 out of 1.25 points

_____ is a form of unsupervised dimensionality reduction.

Selected Answer: ✓ PCA

Answers: ✓ PCA

       LDA

       SVM

       Partial Discriminant Analysis

Question 4

1.25 out of 1.25 points

In practice, feature extraction is NOT used to:

Selected Answer: ✓ preserve the integrity of the original feature set

Answers: improve computational efficiency of the learning algorithm

       ✓ preserve the integrity of the original feature set

       improve the predictive performance by reducing the curse of dimensionality

       improve storage space

Question 5

1.25 out of 1.25 points

In a nutshell, PCA aims to find the directions of maximum variance in high-dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original one.

Selected Answer: ✓ True

Answers: ✓ True

       False

Question 6

1.25 out of 1.25 points

The orthogonal axes (principal components) of the new subspace can be interpreted as the directions of minimum variance given the constraint that the new feature axes are orthogonal to each

other,

Selected Answer: ✅ False
Answers: True
✅ False

Question 7

1.25 out of 1.25 points

The first principal component will have the largest possible variance, and all consequent principal components will have
the largest variance given the constraint that these components are uncorrelated (orthogonal) to the other principal components.

Selected Answer: ✅ True
Answers: ✅ True
False

Question 8

1.25 out of 1.25 points

Even if the input features are correlated, the resulting principal components will be mutually orthogonal (uncorrelated).

Selected Answer: ✅ True
Answers: ✅ True
False

Question 9

1.25 out of 1.25 points

It is not necessary to standardizer the features prior to PCA.

Selected Answer: ✅ False
Answers: True
✅ False

Question 10

1.25 out of 1.25 points

The eigenvalues of the covariance matrix represent the principal components (the directions of maximum variance), whereas the corresponding eigenvectors will define their magnitude.

Selected Answer: ✅ False
Answers: True
✅ False

Question 11

8.75 out of 8.75 points

(Each answer here counts as a separate answer (7 questions total) so take your time and be careful!)

Put the following steps in sequence to correctly perform PCA algorithm:

Answers	Selected Answer
✅ 1. Standardize the d-dimensional dataset.	✅ 1. Standardize the d-dimensional dataset.
✅ 2. Construct the covariance matrix.	✅ 2. Construct the covariance matrix.
✅ 3. Decompose the covariance matrix into its eigenvectors and eigenvalues.	✅ 3. Decompose the covariance matrix into its eigenvectors and eigenvalues.
✅ 4. Sort the eigenvalues by decreasing order to rank the corresponding eigenvectors.	✅ 4. Sort the eigenvalues by decreasing order to rank the corresponding eigenvectors.
✅ 5. Select k eigenvectors which correspond to the k largest eigenvalues, where k is the dimensionality of the new feature subspace (k less than or equal to d).	✅ 5. Select k eigenvectors which correspond to the k largest eigenvalues, where k is the dimensionality of the new feature subspace (k less than or equal to d).
✅ 6. Construct a projection matrix W from the "top" k eigenvectors.	✅ 6. Construct a projection matrix W from the "top" k eigenvectors.
✅ 7. Transform the d-dimensional input dataset X using the projection matrix W to obtain the new k-dimensional feature subspace.	✅ 7. Transform the d-dimensional input dataset X using the projection matrix W to obtain the new k-dimensional feature subspace.

Question 12

1.25 out of 1.25 points

Whereas PCA attempts to find the orthogonal component axes of maximum variance in a dataset, the goal in LDA is to find the feature subspace that optimizes class separability.

Selected Answer: ✅ True

Answers: ✅ True
False

Question 13

1.25 out of 1.25 points

Both PCA and LDA are linear transformation techniques that can be used to reduce the number of dimensions in a dataset; the former is a supervised algorithm, whereas the latter is unsupervised.

Selected Answer: ✅ False

Answers: True
✅ False

Question 14

1.25 out of 1.25 points

We cannot tackle nonlinear problems simply by projecting them onto a new feature space of higher dimensionality in order to make the classes linearly separable.

Selected Answer: ✅ False

Answers: True

✅ False

Sunday, October 21, 2018 12:57:41 PM CDT

← **OK**

Review Test Submission: Quiz 6

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Quiz 6
Started	10/4/18 10:12 PM
Submitted	10/4/18 10:19 PM
Due Date	10/7/18 11:59 PM
Status	Completed
Attempt Score	22.5 out of 25 points
Time Elapsed	6 minutes out of 10 minutes
Instructions	This quiz may contain questions of the following types: True/false, multiple choice, fill in the blank, short answer or matching. It is open book and open notes, but you may not consult with anyone while taking the test. It should be your own work.

It consists of *10* questions worth an equal amount each. It is timed to *10* minutes. It must be completed in one sitting. It will auto-submit at the end of the timed interval automatically. |
| Results Displayed | All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions |

Question 1

0 out of 2.5 points

Using the holdout method, we split our initial dataset into a separate training and test dataset—the former is used
for model training, and the latter is used to estimate its generalization performance.

Selected Answer: ❌ False

Answers: ✅ True

              False

 Response Feedback: This is true.

Question 2

2.5 out of 2.5 points

A better way of using the holdout method for model selection is to separate the data into three parts:
a training set, a validation set, and a test set. The training set is used to fit the different models, and

the performance on the validation set is then used
for the model selection.

Selected Answer: ✅ True
Answers: ✅ True
False

Question 3

2.5 out of 2.5 points

In k-fold cross-validation, we randomly split the training dataset into k folds without replacement, where k — 1 folds are used for the model training, and one fold is used for performance evaluation. This procedure is repeated k times so that we obtain k models and performance estimates.

Selected Answer: ✅ True
Answers: ✅ True
False

Question 4

2.5 out of 2.5 points

A perfect classifier would fall into the _____ corner of the ROC graph with a TPR of 1 and an FPR of 0.

Selected Answer: ✅ top left
Answers: ✅ top left
top right
bottom left
bottom right

Question 5

2.5 out of 2.5 points

Since k-fold cross-validation is a resampling technique without replacement, the advantage of this approach is that each sample point will be used for validation (as part of a test fold) exactly ten times.

Selected Answer: ✅ False
Answers: True
✅ False

Question 6

2.5 out of 2.5 points

In k-fold cross validation the maximum value for k is 10, although smaller values can be used.

Selected Answer: ✅ False
Answers: True
✅ False

Question 7

2.5 out of 2.5 points

There is no practical difference between k-fold cross validation and stratified k-fold cross validation when there is an imbalance between the classes in the dataset.

Selected Answer: ✅ False

Answers: True

✅ False

Question 8

2.5 out of 2.5 points

When using stratified sampling, only the test set will preserve the dataset's class balance proportions, not the training set.

Selected Answer: ✅ False

Answers: True

✅ False

Question 9

2.5 out of 2.5 points

The standard tool for reporting performance of a classifier is called a _____ matrix.

Selected Answer: ✅ confusion

Answers: ✅ confusion

correlation

calibration

covariance

Question 10

2.5 out of 2.5 points

The performance metric known as Precision is calculated by dividing the True Positives by both the sum of the true Positives and the false Negatives.

Selected Answer: ✅ False

Answers: True

✅ False

Sunday, October 21, 2018 12:58:19 PM CDT

← **OK**

Course Modules   Module 7 - Ensembling for Performance Improvement
Review Test Submission: Quiz 7 - ensembling

Review Test Submission: Quiz 7 - ensembling

User	Zefang Zhao
Course	Fall 2018-IE 598-Machine Learning in Fin Lab-Section MLF
Test	Quiz 7 - ensembling
Started	10/14/18 3:03 PM
Submitted	10/14/18 3:18 PM
Due Date	10/14/18 11:59 PM
Status	Completed
Attempt Score	23.75 out of 25 points
Time Elapsed	15 minutes out of 15 minutes
Instructions	This quiz may contain questions of the following types: True/false, multiple choice, fill in the blank, short answer or matching. It is open book and open notes, but you may not consult with anyone while taking the test. It should be your own work.

It consists of ****20**** questions worth an equal amount each. It is timed to ****15**** minutes.
It must be completed in one sitting. It will auto-submit at the end of the timed interval automatically. |
| Results Displayed | All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions |

Question 1

1.25 out of 1.25 points

Bagging is also known as bootstrap aggregating because it uses bootstrap samples drawn form the training set without replacement.

Selected Answer: ✅ False
Answers: True
                  ✅ False

Question 2

1.25 out of 1.25 points

From the datacamp exercise, supply the correct function to extract random forest feature importances:

Selected Answer: ✅ random_forest.feature_importances_

Answers: random_forest.feature_importance()

                  random_forest = RandomForestRegressor()

                  random_forest.fit(train_features, train_targets)

                  ✅ random_forest.feature_importances_

Question 3
1.25 out of 1.25 points

It is worth noting that ensemble learning drastically reduces the computational complexity compared to individual classifiers.

Selected Answer: ✅ False
Answers: True
                  ✅ False

Question 4
1.25 out of 1.25 points

It is considered bad practice to select a model based on the repeated usage of the test set. The estimate of the generalization performance may be over-optimistic.

Selected Answer: ✅ True
Answers: ✅ True
                  False

Question 5
1.25 out of 1.25 points

Each bagged subset of the training set contains a certain portion of duplicates and some of the original samples don't appear in a resampled dataset at all.

Selected Answer: ✅ True
Answers: ✅ True
                  False

Question 6
1.25 out of 1.25 points

In contrast to bagging, the initial formulation of boosting, the algorithm uses random subsets of training samples drawn from the training dataset with replacement.

Selected Answer: ✅ False
Answers: True
                  ✅ False

Question 7
1.25 out of 1.25 points

A decision tree model with the max_depth parameter set to 1 is called a decision tree stump.

Selected Answer: ✅ True

Answers: ✅ True
False

Question 8

1.25 out of 1.25 points

From the datacamp exercise, supply the correct missing parameters:
random_forest = RandomForestRegressor(_____=200, max_____=5, max_____=4, random_____=42)

Selected Answer: ✅ n_estimators, _depth, _features, _state

Answers: estimators, _height, _features, _seed

✅ n_estimators, _depth, _features, _state

n_estimators, _depth, _features, _seed

n_estimators, _depth, _scale, _start

Question 9

1.25 out of 1.25 points

AdaBoost uses the complete training set to train the weak learners where the training samples are reweighted in each iteration to build a strong classifier that learns from the mistakes of the previous weak learners in the ensemble.

Selected Answer: ✅ True

Answers: ✅ True
False

Question 10

1.25 out of 1.25 points

From the datacamp exercise, which of the following is the correct statement to print the in-sample accuracy of a decision tree model?

Selected Answer: ✅ print(decision_tree.score(train_features, train_targets))

Answers: ✅ print(decision_tree.score(train_features, train_targets))

print(decision_tree.score(test_features, train_targets))

print(decision_tree.score(test_features, test_targets))

print(decision_tree.score(train_features, test_targets))

Question 11

1.25 out of 1.25 points

In practice, however, boosting algorithms such as AdaBoost are also known for their high variance, that is, the tendency to overfit the training data.

Selected Answer: ✅ True

Answers: ✅ True
False

Question 12

1.25 out of 1.25 points

From the datacamp exercise, which of the following is the correct way to fit a decision tree model?

Selected Answer: ✅ decision_tree.fit(train_features, train_targets)

Answers: decision_tree.predict(test_features)

decision_tree.fit(train_features, test_targets)

✅ decision_tree.fit(train_features, train_targets)

decision_tree.fit(test_features, test_targets)

Question 13

1.25 out of 1.25 points

Boosting, bagging and majority voting are types of ensemble learning methods.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 14

1.25 out of 1.25 points

The purpose of the test set is merely to report an unbiased estimate of the generalization performance of a classifier.

Selected Answer: ✅ True

Answers: ✅ True

False

Question 15

0 out of 1.25 points

Both logistic regression and k-nearest neighbors algorithms (using the Euclidean distance metric) are scale invariant, in contrast to decision trees.

Selected Answer: ❌ True

Answers: True

✅ False

Response Feedback: (F) decision trees are scale invariant, knn and logreg are not.

Question 16

1.25 out of 1.25 points

Ensembles cannot be constructed using different types of individual learners like knn, logistic regression or decision treees.

Selected Answer: ✅ False

Answers: True

Question 17

1.25 out of 1.25 points

More complex classification tasks and a dataset's high dimensionality can easily lead to _____ in _____ decision trees, and this is where the bagging algorithm can really play to its strengths.

Selected Answer:	✅ overfitting, single
Answers:	underfitting, individual
	underfitting, bagged
	✅ overfitting, single
	overfitting, bagged

Question 18

1.25 out of 1.25 points

From the datacamp exercise this week, which of the following is correct to complete the following statement:

days_of_the week = pd. _____ (amd_df.index.dayofweek, prefix='weekday', drop_first=True)

Selected Answer:	✅ get_dummies
Answers:	make_dummies
	✅ get_dummies
	set_dummies
	add_dummies

Question 19

1.25 out of 1.25 points

Using a random forest, we can measure the feature importance as the _____ computed from all decision trees in the forest, without making any assumptions about whether our data is linearly separable or not.

Selected Answer:	✅ averaged impurity decrease
Answers:	average impurity increase
	✅ averaged impurity decrease
	average entropy increase
	maximum impurity decrease

Question 20

1.25 out of 1.25 points

The key concept behind boosting is to focus on training samples that are hard to classify, that is, to let the weak learners subsequently learn from misclassified training samples to improve the

performance of the ensemble.

Selected Answer: ✅ True
Answers: ✅ True
False

Sunday, October 21, 2018 12:58:55 PM CDT

← **OK**