About this Course

8,947 recent views

Are you interested in predicting future outcomes using your data? This course helps you do just that! Machine learning is the process of developing, testing, and applying predictive algorithms to achieve this goal. Make sure to familiarize yourself with course 3 of this specialization before diving into these machine learning concepts. Building on Course 3, which introduces students to integral supervised machine learning concepts, this course will provide an overview of many additional concepts, techniques, and algorithms in machine learning, from basic classification to decision trees and clustering. By completing this course, you will learn how to apply, test, and interpret machine learning algorithms as alternative methods for addressing your research questions.

SHOW ALL

SKILLS YOU WILL GAIN

Data Analysis | Python Programming | Machine Learning | Exploratory Data Analysis

100% online
Start instantly and learn at your own schedule.

Course 4 of 5 in the
Data Analysis and Interpretation Specialization

Flexible deadlines
Reset deadlines in accordance to your schedule.

Approx. 12 hours to complete
Suggested: 3 hours/week

English
Subtitles: English

About Syllabus Reviews Instructors Enrollment Options FAQ

Syllabus - What you will learn from this course

WEEK

1

5 hours to complete

Decision Trees

In this session, you will learn about decision trees, a type of data mining algorithm that can select from among a large number of variables those and their interactions that are most important in predicting the target or response variable to be explained. Decision trees create segmentations or subgroups in the data, by applying a series of simple rules or criteria over and over again, which choose variable constellations that best predict the target variable.

SHOW ALL

7 videos (Total 40 min), 15 readings, 1 quiz **SEE LESS**

7 videos

What Is Machine Learning? 2m

Machine Learning and the Bias Variance Trade-Off 6m

What Is a Decision Tree? 5m

What is the Process of Growing a Decision Tree? 4m

Building a Decision Tree with SAS 9m

Strengths and Weaknesses of Decision Trees in SAS 4m

Building a Decision Tree with Python 9m

15 readings

Some Guidance for Learners New to the Specialization 10m

SAS or Python - Which to Choose? 10m

Getting Started with SAS 10m

Getting Started with Python 10m

Course Codebooks 10m

Course Data Sets 10m

Uploading Your Own Data to SAS 10m

Data Set for Decision Tree Videos (tree_addhealth.csv) 10m

SAS Code: Decision Trees 10m

CART Paper - Prevention Science 10m

Python Code: Decision Trees 10m

Installing Graphviz and pydotplus 10m

Getting Set up for Assignments 10m

Tumblr Instructions 10m

Assignment Example 10m

WEEK

2

3 hours to complete

Random Forests

In this session, you will learn about random forests, a type of data mining algorithm that can select from among a large number of variables those that are most important in determining the target or response variable to be explained. Unlike decision trees, the results of random forests generalize well to new data.

SHOW ALL

4 videos (Total 25 min), 4 readings, 1 quiz **SEE LESS**

4 videos

What Is A Random Forest and How Is It "Grown"? 4m

Building a Random Forest with SAS 7m

Building a Random Forest with Python 6m

Validation and Cross-Validation 7m

4 readings

SAS code: Random Forests 10m

The HPForest Procedure in SAS 10m

Python Code: Random Forests 10m

Assignment Example 10m

WEEK

3

3 hours to complete

Lasso Regression

Lasso regression analysis is a shrinkage and variable selection method for linear regression models. The goal of lasso regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The lasso does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients variables are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical or both. In this session, you will apply and interpret a lasso regression analysis. You will also develop experience using k-fold cross validation to select the best fitting model and obtain a more accurate estimate of your model's test error rate. To test a lasso regression model, you will need to identify a quantitative response variable from your data set if you haven't already done so, and choose a few additional quantitative and categorical predictor (i.e. explanatory) variables to develop a larger pool of predictors. Having a larger pool of predictors to test will maximize your experience with lasso regression analysis. Remember that lasso regression is a machine learning method, so your choice of additional predictors does not necessarily need to depend on a research hypothesis or theory. Take some chances, and try some new variables. The lasso regression analysis will help you determine which of your predictors are most important. Note also that if you are working with a relatively small data set, you do not need to split your data into training and test data sets. The cross-validation method you apply is designed to eliminate the need to split your data when you have a limited number of observations.

SHOW ALL

5 videos (Total 32 min), 3 readings, 1 quiz **SEE LESS**

5 videos

What is Lasso Regression? 4m

Testing a Lasso Regression with SAS 10m

Data Management for Lasso Regression in Python 3m

Testing a Lasso Regression Model in Python 10m

Lasso Regression Limitations 2m

3 readings

SAS Code: Lasso Regression 10m

Python Code: Lasso Regression 10m

Assignment Example 10m

WEEK

4

3 hours to complete

K-Means Cluster Analysis

Cluster analysis is an unsupervised machine learning method that partitions the observations in a data set into a smaller set of clusters where each observation belongs to only one cluster. The goal of cluster analysis is to group, or cluster, observations into subsets based on their similarity of responses on multiple variables. Clustering variables should be primarily quantitative variables, but binary variables may also be included. In this session, we will show you how to use k-means cluster analysis to identify clusters of observations in your data set. You will gain experience in interpreting cluster analysis results by using graphing methods to help you determine the number of clusters to interpret, and examining clustering variable means to evaluate the cluster profiles. Finally, you will get the opportunity to validate your cluster solution by examining differences between clusters on a variable not included in your cluster analysis. You can use the same variables that you have used in past weeks as clustering variables. If most or all of your previous explanatory variables are categorical, you should identify some additional quantitative clustering variables from your data set. Ideally, most of your clustering variables will be quantitative, although you may also include some binary variables. In addition, you will need to identify a quantitative or binary response variable from your data set that you will not include in your cluster analysis. You will use this variable to validate your clusters by evaluating whether your clusters differ significantly on this response variable using statistical methods, such as analysis of variance or chi-square analysis, which you learned about in Course 2 of the specialization (Data Analysis Tools). Note also that if you are working with a relatively small data set, you do not need to split your data into training and test data sets.

SHOW ALL

6 videos (Total 42 min), 3 readings, 1 quiz SEE LESS

6 videos

What Is a k-Means Cluster Analysis? 6m

Running a k-Means Cluster Analysis in SAS, pt. 1 8m

Running a k-Means Cluster Analysis in SAS, pt. 2 6m

Running a k-Means Cluster Analysis in Python, pt. 1 8m

Running a k-Means Cluster Analysis in Python, pt. 2 10m

k-Means Cluster Analysis Limitations 2m

3 readings

SAS Code: k-Means Cluster Analysis 10m

Python Code: k-Means Cluster Analysis 10m

Assignment Example 10m

4.2 ★★★★½

50 Reviews >

29% started a new career after completing these courses

36% got a tangible career benefit from this course

17% got a pay increase or promotion

Top reviews from Machine Learning for Data Analysis

★★★★★ By BC • OCT 5TH 2016

Very good course. I recommend to anyone who's interested in data analysis and machine learning.

★★★★★ By EM • JUN 26TH 2016

Good introduction with python example for famous algorithm such as random forest and k-mean

Instructors

Jen Rose

Research Professor
Psychology

Lisa Dierker

Professor
Psychology

About Wesleyan University

At Wesleyan, distinguished scholar-teachers work closely with students, taking advantage of fluidity among disciplines to explore the world with a variety of tools.
The university seeks to build a diverse, energetic community of students, faculty,
and staff who think critically and creatively and who value independence of mind
and generosity of spirit.

About the Data Analysis and Interpretation Specialization

Learn SAS or Python programming, expand your knowledge of analytical methods and applications, and conduct original research to inform complex decisions. The Data Analysis and Interpretation Specialization takes you from data novice to data expert in just four project-based courses. You will apply basic data science **... MORE**

Start Learning Today

Financial aid available

- This Course Plus the Full Specialization
- Shareable Certificates
- Self-Paced Learning Option
- Course Videos & Readings
- Practice Quizzes
- Graded Assignments with Peer Feedback
- Graded Quizzes with Feedback
- Graded Programming Assignments

Enroll for Free

Starts Dec 17

34,308 already enrolled!

Shareable on LinkedIn

Machine Learning for Data Analysis

You can share your Course Certificates in the Certifications section of your LinkedIn profile, on printed resumes, CVs, or other documents.

Other courses in this Specialization

Data Managemen...

Wesleyan University

1 COURSE

Data Analysis Tools

Wesleyan University

1 COURSE

Regression Modeling i...

Wesleyan University

1 COURSE

Data Analysis an...

Wesleyan University

1 COURSE

Frequently Asked Questions

When will I have access to the lectures and assignments?

Once you enroll for a Certificate, you'll have access to all videos, quizzes, and programming assignments (if applicable). Peer review assignments can only be submitted and reviewed once your session has begun. If you choose to explore the course without purchasing, you may not be able to access certain assignments.

What will I get if I subscribe to this Specialization?

When you enroll in the course, you get access to all of the courses in the Specialization, and you earn a certificate when you complete the work. Your electronic Certificate will be added to your Accomplishments page - from there, you can print your Certificate or add it to your LinkedIn profile. If you only want to read and view the course content, you can audit the course for free.

What is the refund policy?

If you subscribed, you get a 7-day free trial during which you can cancel at no penalty. After that, we don't give refunds, but you can cancel your subscription at any time. See our full refund policy.

Is financial aid available?

Yes, Coursera provides financial aid to learners who cannot afford the fee. Apply for it by clicking on the Financial Aid link beneath the "Enroll" button on the left. You'll be prompted to complete an application and will be notified if you are approved. You'll need to complete this step for each course in the Specialization, including the Capstone Project. Learn more.

More questions? Visit the Learner Help Center.