



Microsoft: DAT203x Data Science and Machine Learning Essentials

- ▶ Before You Start
- ▶ Module 1: Introduction and Data Science Theory
- ▶ Module 2: Working with Data
- ▶ Module 3: Visualization, and Building and Evaluating Models
- ▼ Module 4: Regression, Classification, and Unsupervised Learning

Chapter 16:
Regression
Modeling

Lab 4A: Working
with Regression
Models

**Chapter 17:
Classification
Modeling**

Lab 4B: Working
with Classification
Models

Chapter 18:
Unsupervised
Learning Models

Lab 4C: Working
with Unsupervised
Learning Models

KEY POINTS


- The process for creating classification models is similar to that for regression models.
- As with regression models, you can use the **Evaluate Model** module to determine model performance, you can use the **Sweep Parameters** module to train a classification module based on a random selection of parameter settings, and you can use the **Cross Validate** module to ensure that the function and coefficients in the model are generalizable.
- The metrics for evaluating classification model performance are based on statistics for *true positive* (TP), *true negative* (TN), *false positive* (FP), and *false negative* (FN) predictions. These metrics include:
 - Accuracy = $TP + TN / (TP + TN + FP + FN)$
 - Precision (or positive predictive value) = $TP / (TP + FP)$
 - Recall = $TP / (TP + FN)$
 - F1 = Precision * Recall / (Precision + Recall)

FURTHER READING

Note: These links take you to external sites outside of the edX course.

- Classification Models: <https://msdn.microsoft.com/en-us/library/azure/dn905808.aspx>

Module 4 Review

Homework due Oct 30,
2015 at 00:00 UTC 

- ▶ Module 5:
Recommenders
and Publishing
Your Work
- ▶ Final Exam

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

