



Like this course? Become an expert by joining the [Machine Learning Specialization \(/specializations/machine-learning\)](https://specializations/machine-learning).

Upgrade

(https://accounts.coursera.org/i/zendesk/courserahelp?return_to=https://learner.coursera.help/hc)



Congratulations!

You've completed the course.

View Final Grade

(</learn/ml-classification/home/welcome>)

Preventing Overfitting in Decision Trees



Out of all machine learning techniques, decision trees are amongst the most prone to overfitting. No practical implementation is possible without including approaches that mitigate this challenge. In this module, through various visualizations and investigations, you will investigate why decision trees suffer from significant overfitting problems. Using the principle of Occam's razor, you will mitigate overfitting by learning simpler trees. At first, you will design algorithms that stop the learning process before the decision trees become overly complex. In an optional segment, you will design a very practical approach that learns an overly-complex tree, and then simplifies it with pruning. Your implementation will investigate the effect of these techniques on mitigating overfitting on our real-world loan data set.

^ Less

Help Center

Overfitting in decision trees



Slides presented in this module

(/learn/ml-classification/supplement/JhRwM/slides-presented-in-this-module)



A review of overfitting 3 min

(/learn/ml-classification/lecture/czRmA/a-review-of-overfitting)



Overfitting in decision trees 6 min

(/learn/ml-classification/lecture/XcPVL/overfitting-in-decision-trees)

Early stopping to avoid overfitting



Principle of Occam's razor: Learning simpler decision trees 5 min

(/learn/ml-classification/lecture/tUvBS/principle-of-occams-razor-learning-simpler-decision-trees)



Early stopping in learning decision trees 7 min

(/learn/ml-classification/lecture/gCuZ8/early-stopping-in-learning-decision-trees)

(OPTIONAL LESSON) Pruning decision trees



(OPTIONAL) Motivating pruning 8 min

(/learn/ml-classification/lecture/9nMdb/optional-motivating-pruning)



(OPTIONAL) Pruning decision trees to avoid overfitting 6 min

(/learn/ml-classification/lecture/qvf6v/optional-pruning-decision-trees-to-avoid-overfitting)



(OPTIONAL) Tree pruning algorithm 4 min

(/learn/ml-classification/lecture/wmODB/optional-tree-pruning-algorithm)

Summarizing preventing overfitting in decision trees



Recap of overfitting and regularization in decision trees 1 min

(/learn/ml-classification/lecture/bRwHo/recap-of-overfitting-and-regularization-in-decision-trees)



Quiz: Preventing Overfitting in Decision Trees 11 questions

(/learn/ml-classification/exam/NDTdj/preventing-overfitting-in-decision-trees)

Programming Assignment



Decision Trees in Practice

(/learn/ml-classification/supplement/AqDoX/decision-trees-in-practice)



Quiz: Decision Trees in Practice 15 questions

(/learn/ml-classification/exam/xRbhG/decision-trees-in-practice)

Handling Missing Data



Real-world machine learning problems are fraught with missing data. That is, very often, some of the inputs are not observed for all data points. This challenge is very significant, happens in most cases, and needs to be addressed carefully to obtain great performance. And, this issue is rarely discussed in machine learning courses. In this module, you will tackle the missing data challenge head on. You will start with the two most basic techniques to convert a dataset with missing data into a clean dataset, namely skipping missing values and imputing missing values. In an advanced section, you will also design a modification of the decision tree learning algorithm that builds decisions about missing data right into the model. You will also explore these techniques in your real-data implementation.

[^ Less](#)

Basic strategies for handling missing data

[Start Lesson](#)

(/learn/ml-classification/lecture/VtYde/challenge-of-missing-data)



Slides presented in this module

(/learn/ml-classification/supplement/zzmdu/slides-presented-in-this-module)



Challenge of missing data 4 min

(/learn/ml-classification/lecture/VtYde/challenge-of-missing-data)



Strategy 1: Purification by skipping missing data 4 min

(/learn/ml-classification/lecture/RUkgm/strategy-1-purification-by-skipping-missing-data)



Strategy 2: Purification by imputing missing data 5 min

(/learn/ml-classification/lecture/4afWj/strategy-2-purification-by-imputing-missing-data)

Strategy 3: Modify learning algorithm to explicitly handle missing data



Modifying decision trees to handle missing data 5 min

(/learn/ml-classification/lecture/rs1r1/modifying-decision-trees-to-handle-missing-data)



Feature split selection with missing data 6 min

(/learn/ml-classification/lecture/ruaji/feature-split-selection-with-missing-data)

Summarizing handling missing data



Recap of handling missing
data 2 min

(/learn/ml-classification/lecture/f6DBX/recap-of-handling-
missing-data)



Quiz: Handling Missing
Data 7 questions

(/learn/ml-classification/exam/fEDCt/handling-missing-
data)