



Microsoft: DAT210x Programming with Python for Data Science



Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▶ 5. Data Modeling
- ▼ **6. Data Modeling II**

Lecture: SVC

Quiz



Lab: SVC

Lab



Lecture: Decision Trees

Quiz

**Lab: Decision Trees**

Lab



6. Data Modeling II > Lab: Decision Trees > Assignment 4



Bookmark

Welcome to Module 6's Decision Tree Labs!

In order to complete the Decision Tree labs in this module, please make sure you download and unarchive this .zip file with all the datasets and files necessary.

Lab Assignment 4

In this lab, you'll get started with decision trees by revisiting UCI's wheat-seeds dataset, so you can benchmark how long it takes to train and predict with decision trees relative to the speed of KNeighbors and SVC, as well as compare the decision boundary plots produced by it.

1. No starter code this time. Instead, take your completed Module6/**assignment1.py** and modify it by adding in a Decision Tree Classifier, setting its max_depth to 9, but not altering any other setting.
2. Make sure you add in the benchmark and drawPlots call for our new classifier as well.
3. Answer the questions below.

Lab Question 1

(1/1 point)

Dive Deeper

To make sure everyone is on the same page, irrespective of the parameters you had on assignment1.py, go ahead and make the following changes:

`C = 1`


`kernel = 'linear'`

`iterations = 5000`

`n_neighbors = 5`

`max_depth = 9`

Which of the following statements is true?

☒ DTrees beat SVC's score in the original, high-D feature space; but SVC has the highest 2D score


☐ KNeighbors has the highest 2D score; but no one beats SVC's score in the original, high-D feature space

☐ KNeighbors and DTrees both tie in their scoring of accuracy in the original, high-D feature space

☐ DTrees are the worst in the original, high-D feature space; but have the best max 2D score

EXPLANATION

KNeighbors Results

5000 Iterations Training Time: 1.95918607712

5000 Iterations Scoring Time: 3.89638900757

High-Dimensionality Score: 83.607

Max 2D Score: 90.164

SVC Results

5000 Iterations Training Time: 4.01298189163

5000 Iterations Scoring Time: 1.68990087509

High-Dimensionality Score: 86.885

Max 2D Score: 93.443

DTree Results

5000 Iterations Training Time: 2.55399107933

5000 Iterations Scoring Time: 1.31068205833

High-Dimensionality Score: 91.803

Max 2D Score: 90.164

You have used 1 of 2 submissions

Lab Question 2

(1/1 point)

Keep dropping the max_depth of the decision tree down until it's high-dimensionality score is **less** than KNeighbors, and then stop.

What is the max_depth value that you hit?

One ▼



Answer: One

EXPLANATION

Decision trees are highly optimal. On this specific dataset, even with as little as two splits, they still out-do KNeighbors, and even fair well against SVC.

max_depth = 1

KNeighbors Results

5000 Iterations Training Time: 1.9534471035

5000 Iterations Scoring Time: 3.8806810379

High-Dimensionality Score: 83.607

Max 2D Score: 90.164

SVC Results

5000 Iterations Training Time: 3.85418701172

5000 Iterations Scoring Time: 1.6913061142

High-Dimensionality Score: 86.885

Max 2D Score: 93.443

DTree Results

5000 Iterations Training Time: 1.70960593224

5000 Iterations Scoring Time: 1.32432198524

High-Dimensionality Score: 68.852

Max 2D Score: 68.852

You have used 1 of 2 submissions

© 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.

POWERED BY
OPENedX

