



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 > Lecture: Decision Trees > Gotchas!



Bookmark

Decision Tree Gotchas!

Decision Trees have some high points which make them desirable for machine learning. They're easy to interpret, by linearizing them into IF .. AND .. AND .. THEN .. blocks. Both training and testing speed are fast. They work with either categorical or continuous features, with or without encoding, and are invariant to feature of scaling. Moreover, if configured properly, they can pretty decently handle irrelevant and noisy features. In a nutshell, decision trees help you decide the worst, best and *expected* classification labels given various scenarios.

While using them, keep in mind that your data still must be *multivariate* linearly separable for classification to work. The decision surfaces of the tree are still flat, even though they can intersect at angles. Also considering how malleable and how many parameters they have, it's easy to get carried away and overfit your tree unless, particularly if you don't fully test against an independent testing set. A good sign of an overfit tree is its very complex structure, and it having branches that reach out erroneously just to correctly label single records, due to their sensitivity to small, local fluctuations in your training data.

Dive Deeper

© 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.

