Microsoft

Logistic Regression & Random Forests

Heather Grandy
Technical Specialist, Azure Data & AI

Agenda

- What is Logistic Regression?
- What is a Random Forest?
- Why/how do I choose one over the other?

Logistic Regression

- Classification algorithm
- Assign observations to a discrete set of classes
- LR returns a probability value that can be mapped to discrete classes
- Types:
  - Binary (True/False)
  - Multi (Cat, Dog, Fish)

Logistic Regression vs. Linear Regression

- Linear Regression: Continuous predictions (numbers in a range)
  - Ex) Predict student test scores on a scale of 0-100
- Logistic Regression: Discrete predictions – can also view probability of the model's classification
  - Ex) Predict whether a student passed or failed

Decision Tree

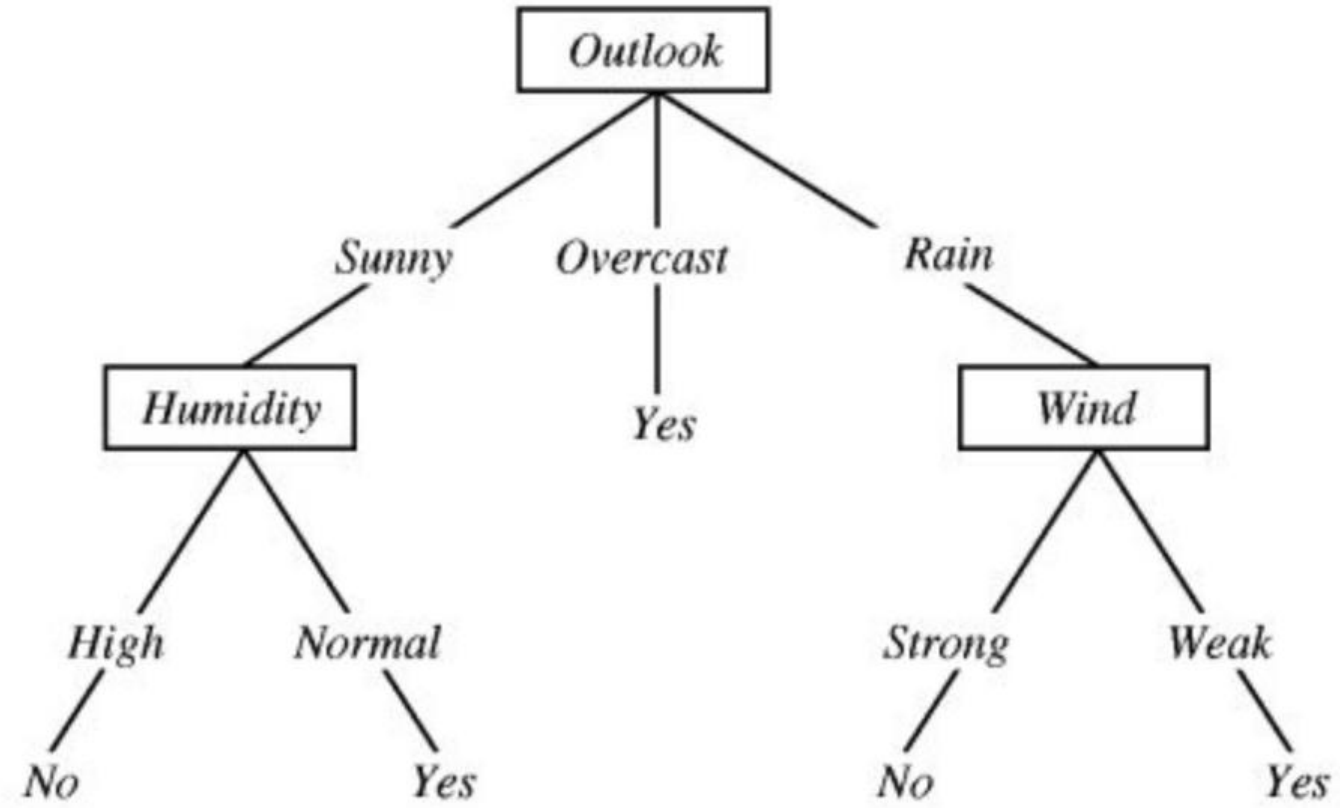
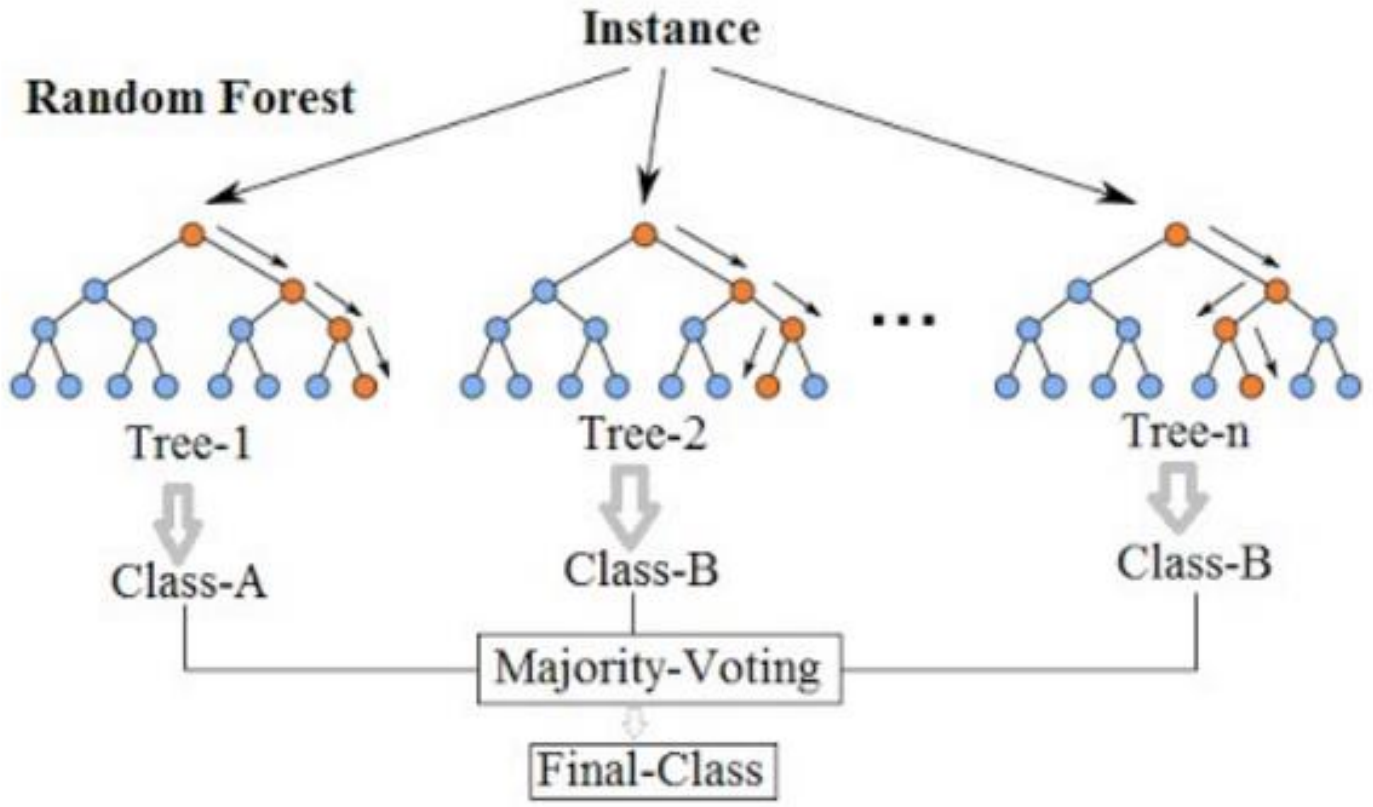


Image courtesy of http://science.slc.edu/~jmarshall/courses/2005/fall/cs151/lectures/decision-trees/

Random Forest

- Classification or regression algorithm
- The random forest algorithm builds several decision trees and then computes the **average** or takes a **majority vote**

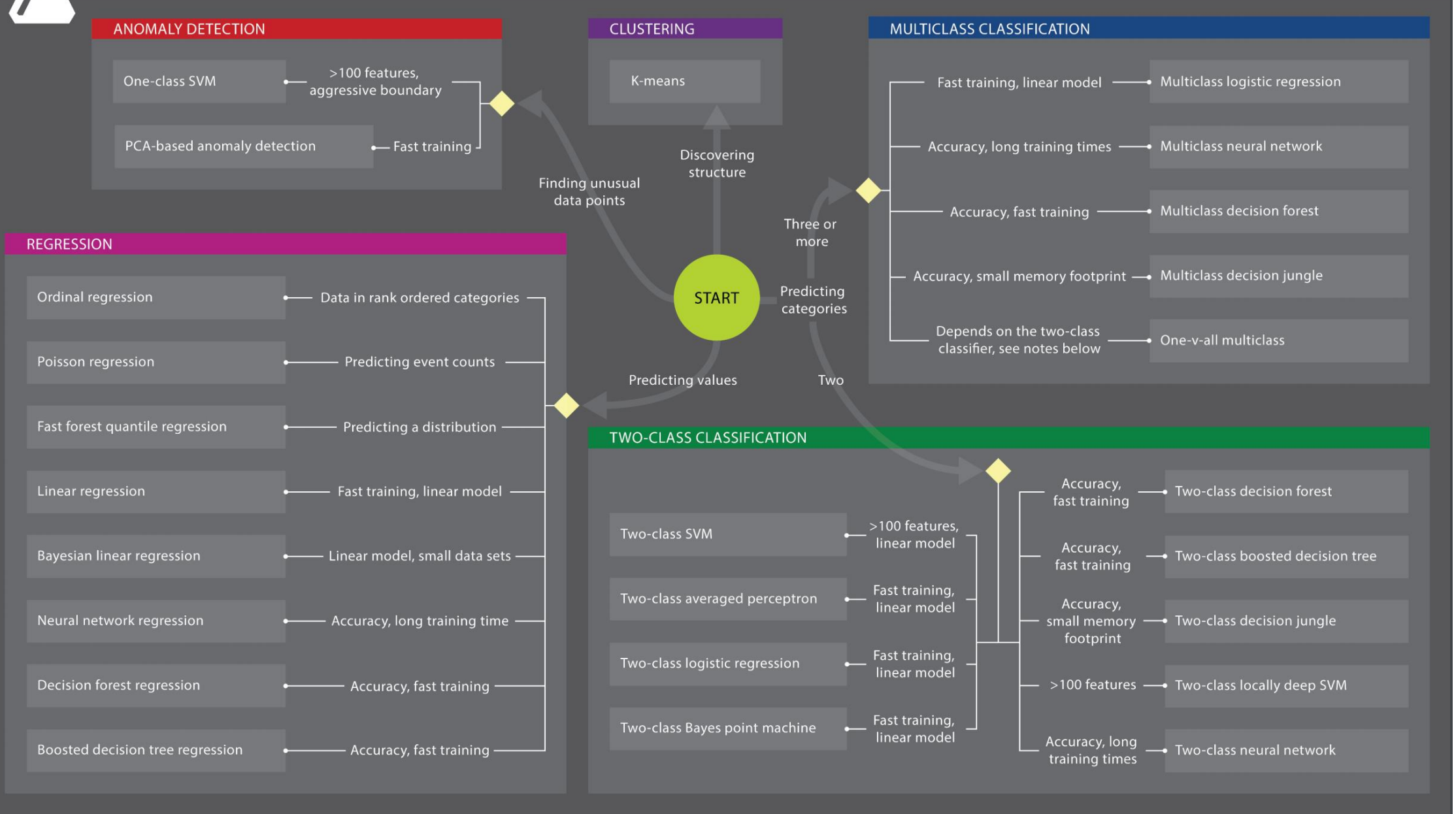


Logistic Regression vs. Random Forests

- Why/how do I choose one over the other? **It depends!**
- Some decision criteria:
  - Is your data categorical? If so, start with random forest first
  - If speed is crucial, try logistic regression first
  - When in doubt, try both!
- Note that the dataset is more important than the model

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



Created by the Azure Machine Learning Team

Email: AzurePoster@microsoft.com

Download this poster: http://aka.ms/MLCheatSheet

Microsoft

https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-cheat-sheet

Microsoft

Thank you!