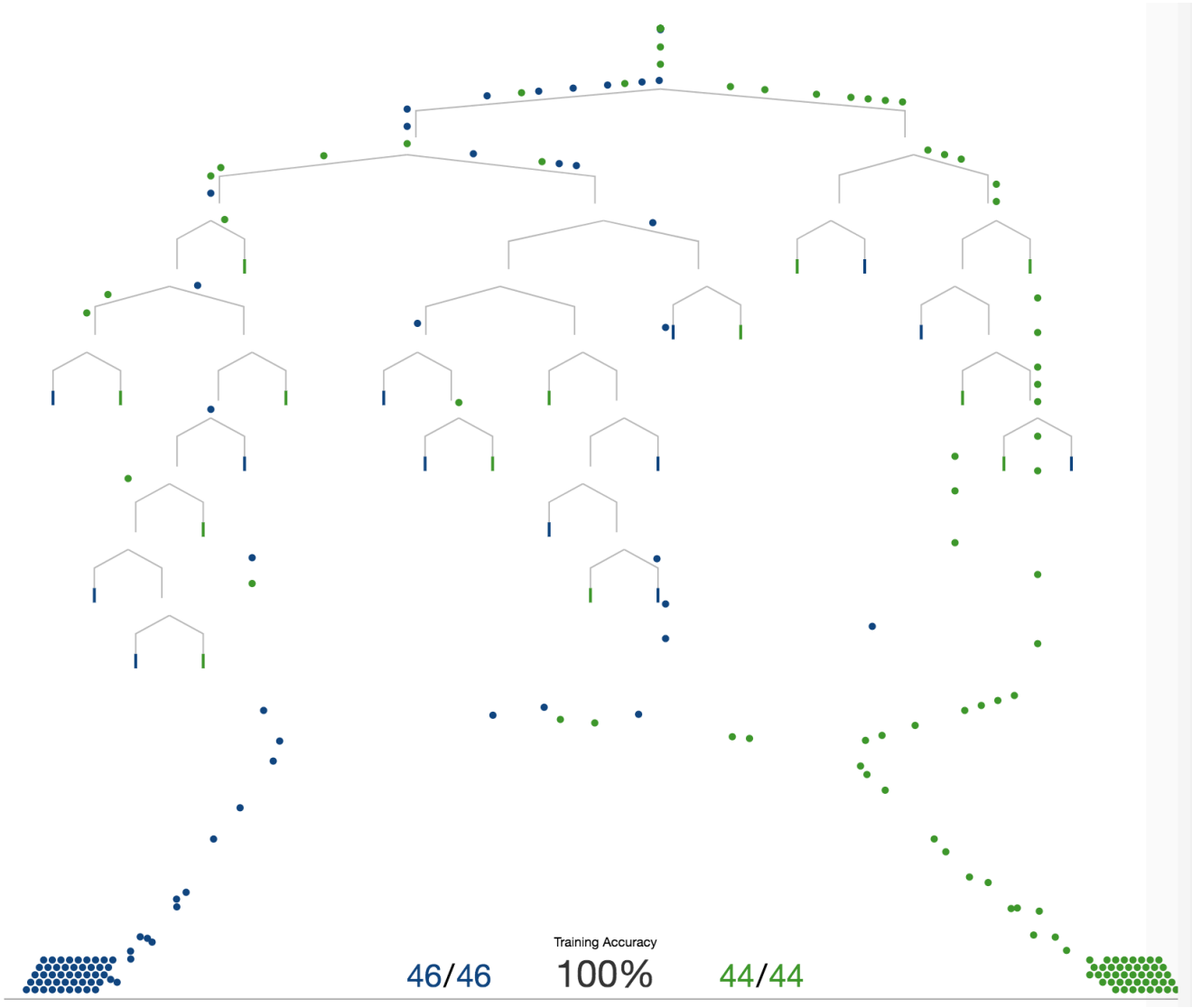
Machine Learning Primer

Michael Schatz

Nov 6, 2019

Lecture 20: Computational Biomedical Research

The newly-trained decision tree model determines whether a home is in San Francisco or New York by running each data point through the branches.



Here you can see the data that was used to train the tree flow through the tree.

This data is called **training data** because it was used to train the model.

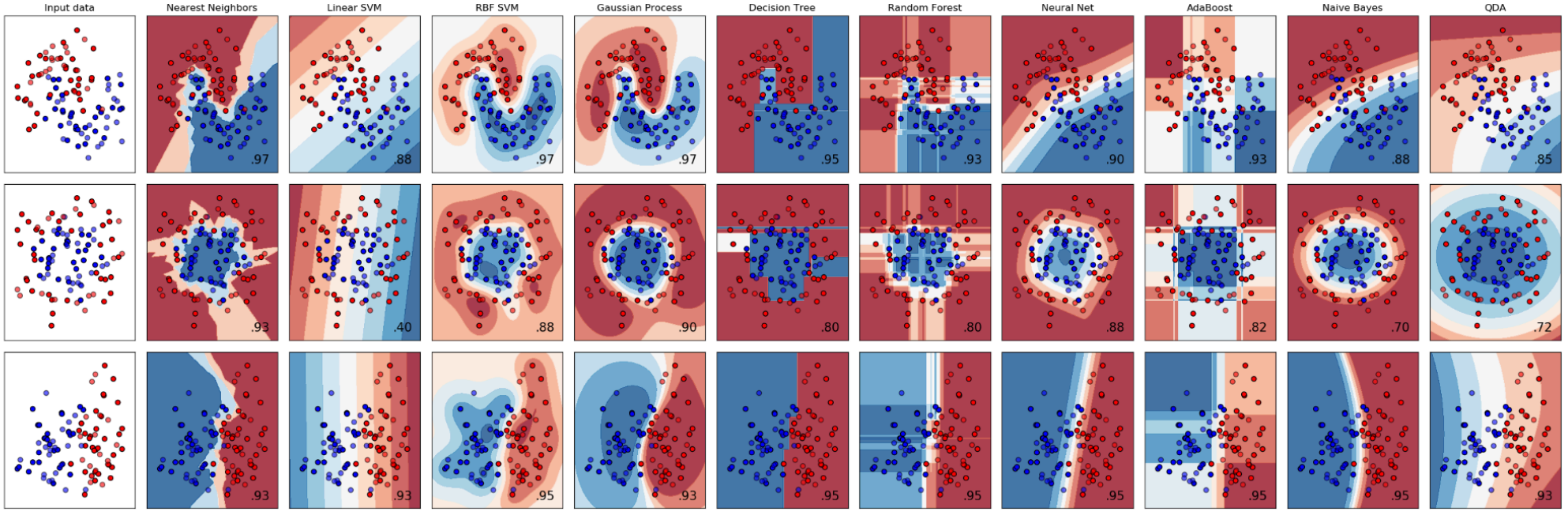
http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

Classifier comparison

A comparison of a several classifiers in scikit-learn on synthetic datasets. The point of this example is to illustrate the nature of decision boundaries of different classifiers. This should be taken with a grain of salt, as the intuition conveyed by these examples does not necessarily carry over to real datasets.

Particularly in high-dimensional spaces, data can more easily be separated linearly and the simplicity of classifiers such as naive Bayes and linear SVMs might lead to better generalization than is achieved by other classifiers.

The plots show training points in solid colors and testing points semi-transparent. The lower right shows the classification accuracy on the test set.



https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html