Ami Ashman

Introduction to AI in Python

2/12/22

Classification Methods

I explored scikit-learn’s documentation and found the following classification methods:

1. Decision trees (DT): DTs are non-parametric supervised learning methods for classification and regression. They create models guided by a collection of conditional rules to predict a target value. The deeper the tree, the more accurate the prediction. DTs work for more multiclass and binary classification.
2. Gaussian Process Classification (GPC): GPCs employ Gaussian processes, a generic supervised learning algorithm useful for probabilistic classification problems, where test predictions are in the form of class probabilities (i.e., 30% chance in Class X, 55% chance in Class Y, 15% chance in Class Z)
3. Nearest Neighbors Classification: This classification paradigm is an example of instance-based learning, meaning it does not try to a general-purpose model, but simply just saves the training data. Scikit-learn has two different types of NNCs:
   1. RadiusNeighborsClassifier, which is based on the number of neighbors within a fixed radius.
   2. KNeighborsClassifier: which is more common and is based on a certain number of neighbors, k, as denoted by the user.
4. Nearest Centroid Classifier: Similar to the previous example, the Nearest Centroid Classifier represents each class as a centroid, hence the name. It classifies new data value depending on which established centroid it falls into. It is a good baseline classifier as there are no parameters to choose or designate.
5. Support Vector Machines (SVM): SVMs can be used for classification, but also for regression and anomaly detection. They are useful for multidimensional data structures and have the added benefits of customization via different Kernel functions.