Overview of Common Machine Learning Models

Intro

This is simply an overview of a few off the most common models used in machine learning, and hopefully this will build interest and a familiarity with the basic concepts of the algorithms. We will cover four models in this post: linear regression, logistic regression, decision trees, and random forests.

Overview

Linear Regression

Linear regression is the most basic machine learning model, and is modeled by the equation

$$y = b + w_0 * x_0 + w_1 * x_1 + w_2 * x_2 + \cdots + w_n * x_n$$

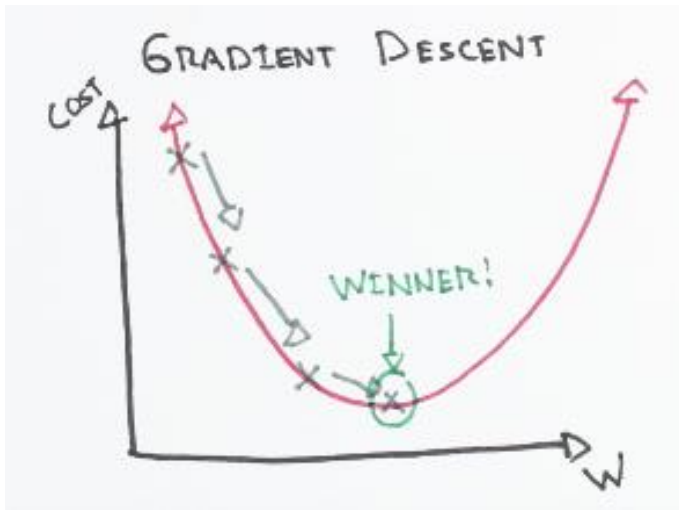
$$where\ x_n\ is\ the\ feature\ column,\ w_n\ is\ the\ weight, and\ b\ is\ the\ intercept$$

The optimal weights are found using the normal equation. It is not covered in this post, but I have linked it here for reference.

Logistic Regression

Logistic regression is used for binary classification problems where the output is a probability between 0 and 1 of the input belonging to the class. It is important to note that logistic regression outputs a *probability* ***not*** a 0 or a 1. In many instances, people will classify probabilities greater than 50% as belonging to the class 1 and less than 50% to the class 0. However, this decision boundary is up to **you.** The decision boundary you choose is dependent on the business objectives and performance metric you are trying to optimize for (precision, recall, or AUC). Logistic regression outputs a probability using the Softmax function. Like linear regression, logistic regression requires that the dataset is clean (absence of missing values) and normalization of data. However, unlike linear regression, which has a closed form equation to find the optimal weights, logistic regression uses an algorithm called gradient descent.

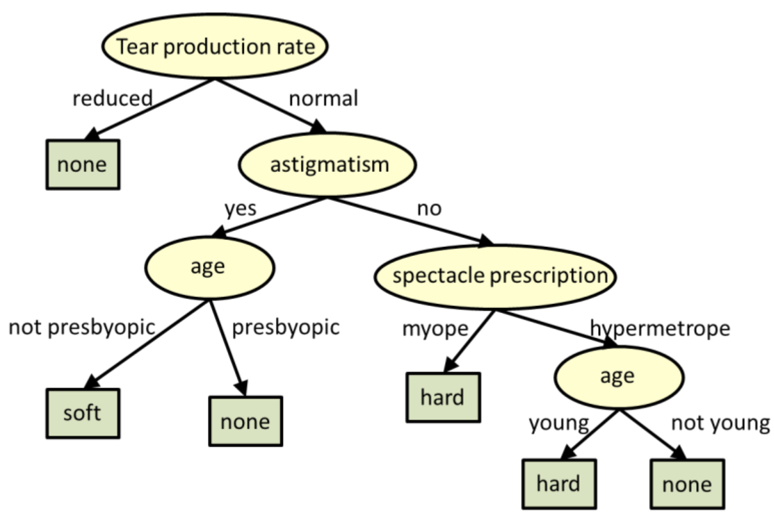
Gradient descent is a very famous algorithm that is the backbone to many machine learning algorithms. Imagine gradient descent to climbing down a hill. The goal is to get to the bottom of the hill, and as you progress downhill you make various adjustments to descend quicker but also safer. During the descend, you learn from mistakes you made previously and adjust course. Similarly, gradient descent works the same way for machine learning models. For machine learning models, the prize at the bottom of the hill is minimal error by achieving the optimal weights. In short, gradient descent is the iterative process the model uses to update weights, make a guess, compare the prediction against the ground truth, and finally make adjustments to adjust course downhill. Once the algorithm reaches the bottom of the gradient, the error is minimized, and the optimal weights are reached. It is important to note; however, that error is usually never 0. If it is, it is more than likely you have overfit the algorithm to the data.



A visualization of gradient descent

Decision Trees

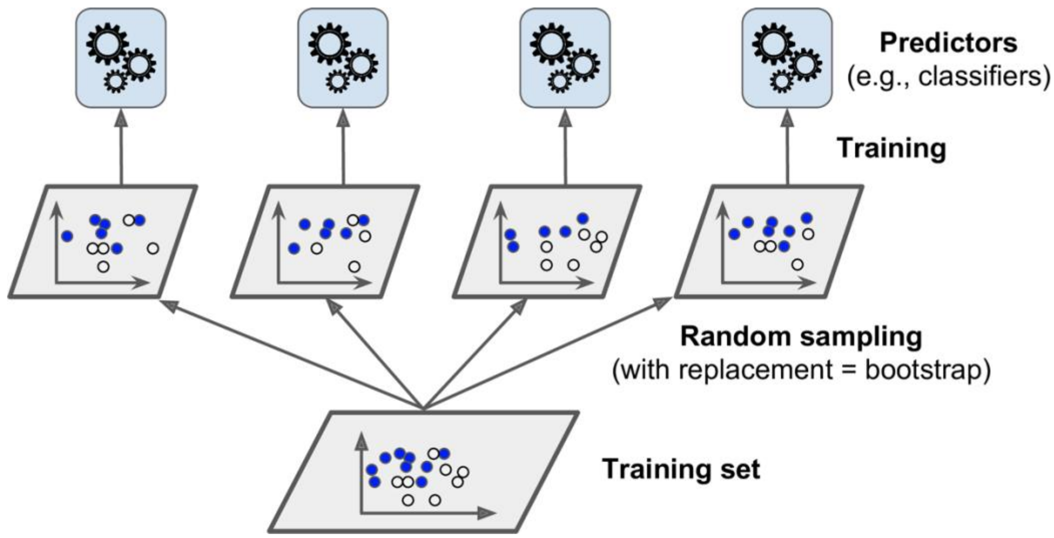
Decision trees are robust machine learning algorithms that are used for both regression *and* classification. In addition, decision trees do not require any data preparation like linear or logistic regression. Decision trees are layers of conditional branches strung together that lead to an outcome. This simplicity combined with their robustness have made decision trees famous. In fact, the choices a decision tree makes can easily be visualized making it easy to identify areas of improvement.



A decision tree classifying tear production based on several factors

Random Forests

Random Forests are a collection of decision trees that are trained on random subsets of the training data. The decision is reached by taking the mode of the predictions in classification problems or the average in regression problems. Each individual decision tree has higher bias, but the aggregation of decision trees reduces both bias and variance. The aggregation of several decision trees reduces both under of the collection of the decision tree is made



Individual predictors, decision trees, trained on random samples with replacement of the training data