Apply machine learning to your capstone project

Max Davis

10/2/2019

**How do you frame your main question as a machine learning problem? Is it a supervised or unsupervised problem? If it is supervised, is it a regression or a classification?**

The main question of this capstone is whether a case of a wind turbine being “down”, that is, producing below 50 kW of power while wind speeds are above a 5 m/s threshold ­–—can be predicted based on the available data. As a machine learning problem, this is a matter of defining a feature that reflects the conditions we are trying to learn about, and figuring out which methods best predict it. Since this is a case of “is down” or “is not down”, it is a binary classification problem. With this large dataset that can be split into test and training sets, it becomes a case of supervised machine learning. The algorithm I have learned to approach binary classification problems is logistic regression.

**What are the main features (also called independent variables or predictors) that you'll use?**

The features that will be investigated as predictors or independent variables in this model, will be all available sensor readings, which are really an average reading for each time stamp in the data. Features with a high proportion of missing values will be removed, narrowing the dataset to 22 independent variables, each a sensor reading average for a given time-stamp.

**Which machine learning technique will you use?**

Classification, specifically logistic regression.

**How will you evaluate the success of your machine learning technique? What metric will you use?**

I will assess the accuracy of the model by using it to predict “down” instances in a test set that was split from a training set. The accuracy, or proportion of acccurately predicted positives and negatives as shown by a confusion matrix will be the measurement of accuracy.