Assignment 4: Machine Learning

INF350E/INF385T Personal Informatics

March 23, 2018

In this assignment, you will explore the fields of applied signal processing and machine learning (ML). You will work on two problems. In the first problem, which is a classical ML-type problem, you will build a classifier that predicts wine quality from its physico-chemical characteristics; in the second problem, which is central to Personal Informatics, you will build a step counter from raw inertial sensor data. Datasets will be provided for both of these problems. Once you complete the assignment, you won't be a ML expert yet but you will have a good sense of the workflow involved in applying ML to solve real, practical problems.

1   Setup

For this assignment, at a minimum, you will need a computer with Python and scikit-learn. You are welcome to use other packages as well, such as pandas. I recommend you install the Anaconda distribution. Pointers to these resources are located at the end of this document.

2   Predicting Wine Quality

The file "winequality-white.csv" is a dataset that includes a long list of wine samples (one per line). For each sample, various physico-chemical characteristics are specified (each characteristic is a feature, structured as its own column in the csv file). Additionally, every wine sample also includes a quality rating produced by a wine expert. Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

You mission, should you choose to accept it, is to use this data to build a classifier using scikit-learn that predicts the quality of a wine sample based on its physico-chemical specifications. You can use any learning algorithm available in scikit-learn to train your classifier, and you should evaluate it using the 10-fold cross-validation technique.

You deliverable should be a python script with the code you used to train the classifier and evaluate it. Hint: The scikit-learn package is very well documented.

3 Step Counter

Prof. Thomaz has recently lost his Fitbit and would like you to build him a new smart step counter based on the raw inertial sensor signals of his smartphone. To help you, he provided a file called "steps.csv"; It contains accelerometer and gyroscope data that he collected at 30Hz while walking and running with his smartphone in his right hand. The csv file contain headers that indicate what each column refers to (e.g., accelerometer X axis, gyroscope Z axis).

Your deliverable for this problem should be a python script that outputs the number of steps recorded in the "steps.csv" file. You are free to solve this problem in any way you want; An approach could be segmenting the signal, extracting frames and features from the signal, and using them to build a classifier to recognize and then count steps. Another approach might be to discover features in the signal that, on their own, are predictive of steps.

To see how your approach generalizes, your code will be evaluated not only on the "steps.csv" file collected by Prof. Thomaz but also on walking/running data of another yet-to-be-determined person. Feel free to test it out with your own walking-running data to see how it performs. To do this you will need to collect some data using one of the apps you downloaded for the motion tracking assignment.

4 Submission Instructions

You will submit this assignment electronically through Canvas. Include your python scripts in a directory named "your-last-name-assignment-4" and zip it. Give the zip file the same name as the directory.

5 Resources

1. http://scikit-learn.org/stable/index.html

2. http://scikit-learn.org/stable/supervised_learning.html

3. http://scikit-learn.org/stable/modules/preprocessing.html

4. https://pandas.pydata.org

5. https://www.anaconda.com/download/