ML Project Notes – OBD Data Analysis

I resume this project after some months of break.

First thing I’ll do is to create an artificial dataset in order to have clearly separable data. For now, I give up the idea of recognizing who is driving. I’m going to try to detect if a driving session is related to an itinerary of type:

* Urban
* (Out-of-town)
* Highway
* Combined

Let’s start from a two-class classification:

City:

* Lower speeds
* More variable speed (traffic)
* More gear changes
* More acceleration (from a standing still up to speed limit)
* More braking (often to a dead stop at intersections, traffic lights and stop signs)

Highway:

* Higher speeds
* More constant speeds
* Less gear changes
* Less acceleration
* Less braking

So, I could measure driving session of about 5 minutes. I’ll have two timeseries: RPM and speed. I can extract the following features:

* Mean speed <v>, maximum speed V, minimum speed v;
* Mean acceleration <a>, max acceleration A, min acceleration a (in module)
* Number of gear changes #GC

It could be good to have also data about throttle position and brakes, but firstly I should find out if my car supports them, secondly, they would increase the sampling time.

NOTE – I put all the files related to the driver recognition in a dedicated folder Old\_DriverRecognition. These files are:

* *main\_OldDriverRecognition*
* *OrganizeCollectedData\_OldDriverRecognition*

In the *main* script I try now to extract the features. I still use the *ComputeFirstDerivative* written some months ago, but I think I should revise it. In particular, in *ComputeGearShifts* I don’t obtain beautiful results: it often happens that a decrease in the number of RPMs is so long that it counts for more than one gear shifts. For now, I continue with this behaviour in order to arrive at the design of the classifier.

So, I can create my dataset. I have 7 features, so it’s seven-dimensional.

The first trial I do is with the *kernel regularized least squares* classifier. My dataset is very small and so the test is not so significant, but it can give me an idea about what’s happening. With 9 driving sessions, three of which are dummy, the error on the little test set is zero. I absolutely have to collect more and more data. In the meantime, I try the other algorithms of classification. I’ll also have to do some cross-validation, remember it!

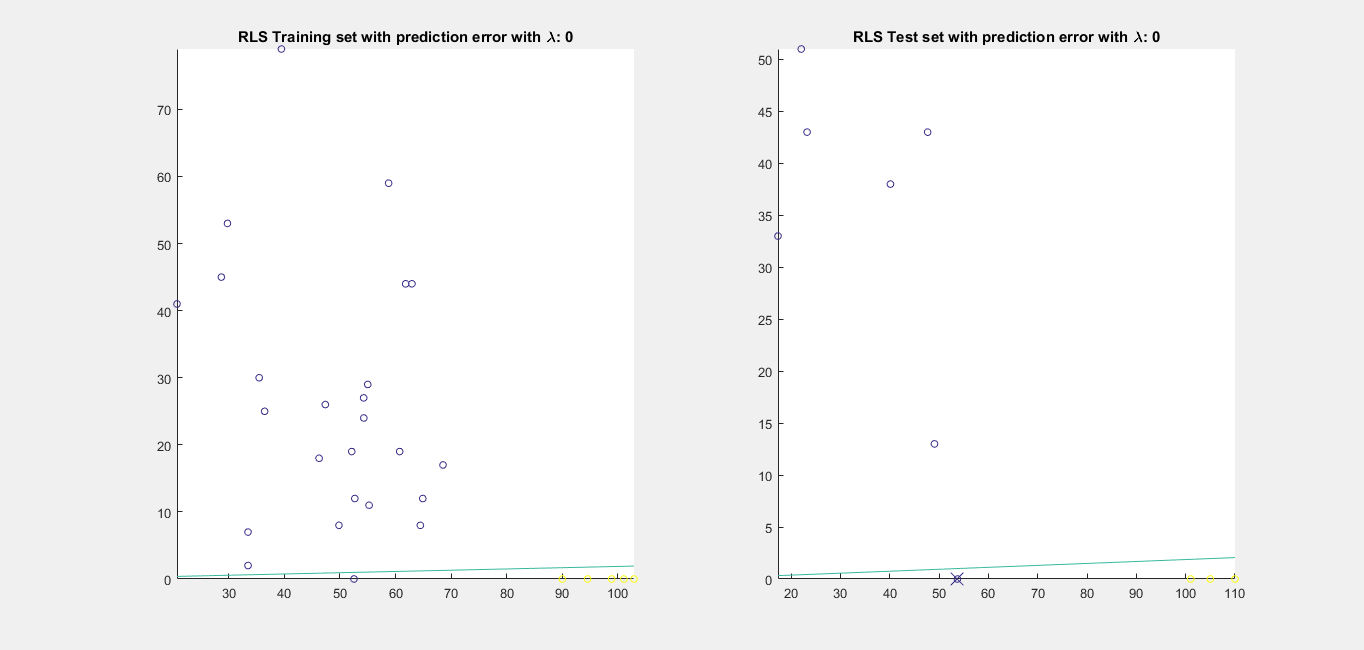
*13/05*

I add some examples to my dataset, some are taken on highway drive, other from out-of-town drive. One of the most interesting example is the one recorder on “sopraelevata”: here there aren’t gear shifts, but the speed is almost constant; I label it as out-of-town and I’m curious to know how the classifier will manage it.

14/06

My dataset is getting a decent size. I begin to collect some screenshot about interesting issues.

Image “RLS\_BinaryClassification\_OneError\_NoOffset”:



Here I can see that the classifier cannot avoid to do an error already in the training set: the hypothesis is linear and it must pass through the origin. There is an example which belongs to “out-of-town”, but it’s probably taken on *sopraelevata*, so the number of gear changes is 0 and there doesn’t exist a linear function that can separate it from other point with zero gear changes. This suggest me to use a linear function with offset.