Machine Learning 2015: Project 2 - Regression Report

pungast@student.ethz.ch

November 20, 2015

Experimental Protocol

I added nonlinear features to the original features and used a Random Forest classifier to predict the 3 classes.

1 Tools

Which tools and libraries have you used (e.g. Matlab, Python with scikit-learn, Java with Weka, SPSS, language x with library y, ...). If you have source-code (Matlab scripts, Python scripts, Java source folder, ...), make sure to submit it on the project website together with this report. If you only used command-line or GUI-tools describe what you did.

2 Algorithm

Describe the algorithm you used for regression (e.g. ordinary least squares, ridge regression, ...)

3 Features

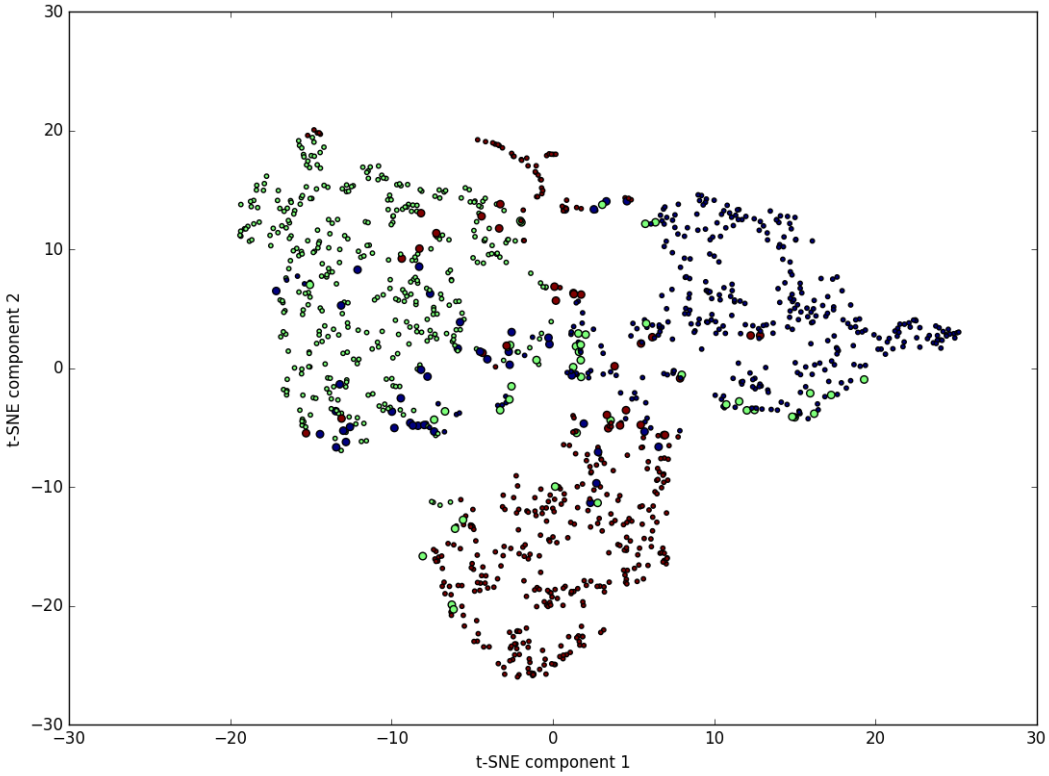
I added the following new features, for each feature x: $logx$, $xlogx$, $\sqrt{x}$, x^3, x^4 and all second-order polynomial combinations of features.

I did not normalise the data because it turned out that neither row-max normalising (dividing all features of each datapoint by the maximum dimension) or column-max normalising (diving all i-th feature values with the maximum i-th feature over all datapoints) improved prediction accuracy in cross-validation.

Figure 3

4 Parameters

How did you find the parameters of your model? (What parameters have you searched over, cross validation procedure, ...)



5 Lessons Learned

What other algorithms, tools or methods did you try out that didn't work well? Why do you think they performed worse than what you used for your final submission?