Human Activity Recognition using
mobile sensors report

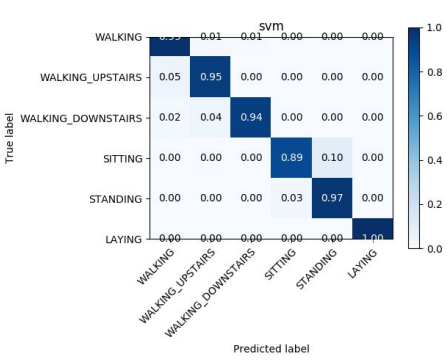
Part 1: Description of the results.

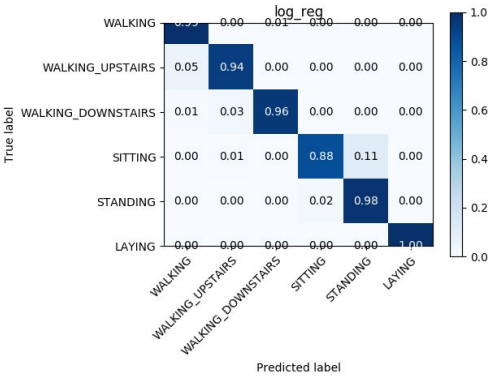
Models with best parameters are evaluated using:
1. Accuracy, Precision, Recall.
2. Confusion Matrix.

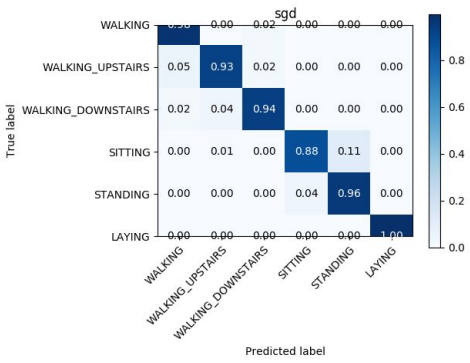
The accuracy outputs (of the last run) are the following:
Sgd: 94.842
Log_reg: 96.098
Svm: 95.792

As we can see, the accuracy of the linear models is quite good. However, we should remember that we have multiple labels in our data, that means that regardless of the high level of accuracy these models can show bad results within a particular class.

For a more accurate score we need to take a look at how well the classifier does for each label individually.
It is possible to find these values on the diagonal of confusion matrix.
The plots are the the following:
The lost (unseen values because of some problems with plotting) on these plots are:
Sgd: 0.98185484
Log_reg: 0.99395161
Svm: 0.9858871







By the confusion matrix it is possible to see that accuracy is still quite good, however all the models have some problems with guessing “sitting” correctly (compared to other classes).

The confusion matrix is also useful for quickly calculating precision and recall.

It is understood that the higher the precision and recall, the better. However, usually it is not possible to achieve maximum precision and recall at the same time.
Therefore, I would like to have some kind of metric that combines information on the accuracy and completeness of the algorithm. So, I decided to use F-score in report also. For these I use classification report

The classification report is following:
Sgd

	precision	recall	f1-score	support
1.0	0.94	0.98	0.96	496
2.0	0.95	0.93	0.94	471
3.0	0.95	0.94	0.95	420
4.0	0.95	0.88	0.92	491
5.0	0.90	0.96	0.93	532
6.0	1.00	1.00	1.00	537

Log_reg

	precision	recall	f1-score	support
1.0	0.94	0.99	0.97	496
2.0	0.97	0.94	0.96	471
3.0	0.99	0.96	0.98	420
4.0	0.98	0.88	0.93	491
5.0	0.90	0.98	0.94	532
6.0	1.00	1.00	1.00	537

Svm

	precision	recall	f1-score	support
1.0	0.94	0.99	0.96	496
2.0	0.95	0.95	0.95	471
3.0	0.99	0.94	0.96	420
4.0	0.96	0.89	0.93	491
5.0	0.91	0.97	0.94	532
6.0	1.00	1.00	1.00	537

Part 2. Best model

Logistic regression and SVM show nearly the same accuracy scores (one can become better than the other one from run to run). Also we can see (by the confusion matrix) that both of them show quite good results within every class.

However, (from part 3) SVM is faster. So, I suppose, it is the best.

Part 3. Fastest model

Time(last run):

sgd time : 67.58840203285217
log_reg time : 264.45946621894836
svm time : 64.55559468269348

SGD and SVM are the fastest two, their results are pretty close to each other.
During the current run, SVM appears to be the best.

4. Feasibility of precision and recall

It possible to compute Precision and Recall for the Multi-Class Problem.
Actually, they are shown in the classification report.

To compute the precision and recall we should:

1) Generate a confusion matrix
    - Diagonals contain the true positives for each label.
    - The sum of a row would be total number of instances predicted as a particular label X
    - The sum of a column would be total number of instances that should have label X
2) The precision of a label x is computed as:
    true positives x/ total number of instances predicted as a particular label X
   The recall of a label x is computed as:
    true positives x/total number of instances that should have label X