Support Vector Machines – Datasets

Emanuel Zaymus

In this report we are going to look at other datasets and accuracy which we can achieve with SVMs.

Datasets:

- ADL Normal Activities (1 Kyoto)
- ADL Activities with Errors (2 Kyoto)
- Daily Life 2010-2011 (17 Aruba)

I will test them with two different configurations of feature vectors:

I. Without CLASS PROBABILITIES OF PREVIOUS VECTOR feature

1. SECONDS FROM MIDNIGHT – FIRST RECORD in the window
2. SECONDS FROM MIDNIGHT – LAST RECORD in the window
3. DAY OF THE WEEK – binary feature
4. SECONDS ELAPSED – between the last and the first record of the window
5. SIMPLE COUNTS OF THE SENSORS

II. With CLASS PROBABILITIES OF PREVIOUS VECTOR feature - `predict_proba` function

1. SECONDS FROM MIDNIGHT – FIRST RECORD in the window
2. SECONDS FROM MIDNIGHT – LAST RECORD in the window
3. DAY OF THE WEEK – binary feature
4. SECONDS ELAPSED – between the last and the first record of the window
5. SIMPLE COUNTS OF THE SENSORS
6. CLASS PROBABILITIES OF PREVIOUS VECTOR

Scaling

`sklearn.preprocessing.`**`StandardScaler`**

`sklearn.preprocessing.`**`RobustScaler`**

Classifier

`sklearn.svm.`**`SVC`**`(kernel=’rbf’, C=1.0, gamma=’scale’)` – feature configuration **I.**

`sklearn.svm.`**`SVC`**`(probability=`**`True`**`, break_ties=`**`True`**`)` – feature configuration **II.**

Testing

`KFold``(n_splits=5, shuffle=True, random_state=0)` – feature configuration **I.**

`KFold``(n_splits=5, shuffle=False)` – feature configuration **II.**

ADL Normal Activities (1 Kyoto)

This is only reminder of results from previous reports.

Table 1 – ADL Normal Activities – Accuracy scores (%)

Window size	CLASS PROBABILITIES OF PREVIOUS VECTOR feature included					
	No			Yes		
	No scaling	Standard Scaler	Robust Scaler	No scaling	Standard Scaler	Robust Scaler
5	35.5507	76.1098	76.8366	35.5512	53.7848	36.3574
7	35.8224	80.9266	80.4650	35.8224	54.0193	42.4312
10	36.2378	84.6190	84.2324	36.2383	55.6939	50.3802
12	36.5201	87.2751	86.5446	36.5202	54.8276	53.5001
15	36.9516	89.8176	89.0786	36.9525	57.2195	54.6587
17	37.2455	91.2101	90.4652	37.2456	58.2178	59.2312
19	37.5438	92.3576	91.9238	37.5440	57.6190	56.7537
22	38.0003	93.0757	93.1939	38.0005	58.3847	55.9898
25	38.4681	93.8282	94.0676	38.4687	58.8321	64.5595
27	38.7863	94.1906	94.3113	38.7865	60.5923	65.4387
30	39.2734	94.3620	94.4842	39.2744	64.1841	70.0660
32	39.6057	94.6313	95.0360	39.6058	64.6542	71.5725
35	40.1139	94.8474	95.3824	40.1146	63.9351	74.8267
37	40.4606	95.0191	95.6125	**40.4605**	**64.9523**	76.1739
40	**40.9909**	**95.0810**	**95.9738**	40.9915	62.4357	**77.6473**

ADL Activities with Errors (2 Kyoto)

Table 2 – ADL Activities with Errors – Accuracy scores (%)

Window size	CLASS PROBABILITIES OF PREVIOUS VECTOR feature included					
	No			Yes		
	No scaling	Standard Scaler	Robust Scaler	No scaling	Standard Scaler	Robust Scaler
5	46.0538	82.1693	82.6022	46.0560	64.7928	57.4475
7	46.4193	85.0223	86.1729	46.4215	64.8703	60.0904
10	46.9788	89.6205	90.2427	46.9806	64.7884	68.8815
12	47.3587	91.3376	91.8031	47.3609	64.9283	71.3831
15	47.9406	93.4643	93.3618	47.9431	65.3982	75.4783
17	48.3369	93.7616	94.4021	48.3392	**65.4221**	76.3088
19	48.7400	94.2720	94.2092	48.7419	65.0716	77.0692
22	49.3566	94.3681	94.5368	49.3587	63.8495	78.4030
25	49.9897	94.6808	94.9798	49.9914	63.8776	**78.8956**
27	50.4201	94.4191	94.9146	50.4222	63.4802	78.1965
30	51.0803	94.7610	**95.1539**	51.0826	64.1371	77.6060
32	51.5305	**94.7810**	94.8912	51.5325	63.8874	77.2106
35	52.2203	94.6665	94.8897	52.2225	63.3592	77.1511
37	52.6906	94.5281	94.8661	52.6928	63.5471	77.0579
40	**53.4117**	94.6584	**95.0010**	**53.4144**	62.6601	76.3101

In the dataset with errors the accuracy dropped little bit for scaled features without the feature CLASS PROBABILITIES OF PREVIOUS VECTOR. For data with the extra feature the success rate increased.

Daily Life 2010-2011 (17 Aruba)

This dataset is really large. For the simplicity I used only 5 days (35 624 sensor events) from it to test the SVM. Anyway, it took quite a bit of time.

Table 3 – Daily life 2010-2011 (5 days) – Accuracy scores (%)

Window size	CLASS PROBABILITIES OF PREVIOUS VECTOR feature included					
	No			Yes		
	No scaling	Standard Scaler	Robust Scaler	No scaling	Standard Scaler	Robust Scaler
5	35.5507	76.1098	76.8366	35.5512	53.7848	36.3574
7	35.8224	80.9266	80.4650	35.8224	54.0193	42.4312
10	36.2378	84.6190	84.2324	36.2383	55.6939	50.3802
12	36.5201	87.2751	86.5446	36.5202	54.8276	53.5001
15	36.9516	89.8176	89.0786	36.9525	57.2195	54.6587
17	37.2455	91.2101	90.4652	37.2456	58.2178	59.2312
19	37.5438	92.3576	91.9238	37.5440	57.6190	56.7537
22	38.0003	93.0757	93.1939	38.0005	58.3847	55.9898
25	38.4681	93.8282	94.0676	38.4687	58.8321	64.5595
27	38.7863	94.1906	94.3113	38.7865	60.5923	65.4387
30	39.2734	94.3620	94.4842	39.2744	64.1841	70.0660
32	39.6057	94.6313	95.0360	39.6058	64.6542	71.5725
35	40.1139	94.8474	95.3824	40.1146	63.9351	74.8267
37	40.4606	95.0191	95.6125	**40.4605**	**64.9523**	76.1739
40	**40.9909**	**95.0810**	**95.9738**	40.9915	62.4357	**77.6473**

Conclusion

As we could see, by adding CLASS PROBABILITIES OF PREVIOUS VECTOR feature we were not able to raise the accuracy. It seems that this feature decreases the success rate significantly.

Github

https://github.com/emanuelzaymus/ActivityRecognition