Instituto Superior Técnico - UL

Speech Processing

Lab3

Autores:	*IST ID:*
Gonçalo Mestre	87005
Ricardo Antão	87107

Professor Maria Isabel Trancoso

---

10 of May of 2020

Contents

1 Part1 **1**

1.1 Question 1 . . . 1

1.2 Question 2 . . . 1

1.3 Question 3 . . . 1

1.4 Question 4 . . . 2

2 Part2 **2**

2.1 Question 1 . . . 2

2.2 Question 2 . . . 2

2.3 Question 3 . . . 3

2.4 Question 4 . . . 3

2.5 Question 5 . . . 3

2.6 Question 6 . . . 4

3 Part3 **4**

3.1 Question 1 . . . 4

3.2 Question 2 . . . 4

3.2.1 One Layer . . . 4

3.2.2 Adding Layers . . . 4

3.2.3 Adding Batch Normalization and Dropout . . . 4

3.3 Question 3 . . . 4

3.3.1 eGemaps . . . 4

3.3.2 IS11 . . . 5

3.3.3 AVEC2013 . . . 5

3.4 Question 4 . . . 6

4 Additional Feature sets **6**

5 Part4 **6**

1 Part1

1.1 Question 1

The files are saved after running the python script in the folder "csvfiles", this folder is created by the python script.

1.2 Question 2

	Speaker		SL		NSL		Average Duration(s)	Total Duration (h:m:s)
	M	F	M	F	M	F		
Train	706	1991	247	825	459	1166	9.071	6:47:46.3
Dev	372	967	194	294	178	673	7.85962	2:55:24
Test	-	-	-	-	-	-	10.2038	3:31:13.1

It is possible to conclude that the training set has a lot more data than the development set which is comprehensible, and that the total duration in time of the training set is just a little higher than the one of the development set.

Regarding the distribution it is possible to clearly see that the data set is unbalanced, since just 39.8% and 36.5% of the data is of sleepy recordings, for the training and development data sets, respectively.

It is not only on the label that the data set is unbalanced, but also the gender of the data set is clearly unbalanced, since 26.3% and 25.3% of the data is of a Male speaker, for the training and development data sets, respectively.

1.3 Question 3

- eGeMaps
  Number of features - 88
  Type of features (Extended GeMAPS)-
  - Temporal Features - 7 features
  - Functional - 81 features

- IS11

  Number of features - 4368

  Type of features -

  - - 3996 features
  - - 360 features

1.4 Question 4

All the csv files can be found on the folder "csvfiles" that is created by the python script part1.py.

2 Part2

2.1 Question 1

- Precision: $\frac{TP}{TP+FP}$
- Recall: $\frac{TP}{TP+FN}$
- F1: $\frac{TP}{TP+\frac{FN+FP}{2}}$
- Accuracy: $\frac{TP+TN}{TP+FP+FN+TN}$

2.2 Question 2

If the classifier only learns the prior distribution of the data in the training set, it will just classify everything as non sleepy since there is an unbalance through the non sleepy side. Therefore, since on the dev partition there are 1339 speakers, in which, 851 non sleepy speakers and 488 sleepy speakers, the metrics obtained for the dev partition will be given below:

- Precision - Class 1(Sleepy) as Positive: $\frac{0}{0+0} = \frac{0}{0} = error$
- Recall - Class 1(Sleepy) as Positive: $\frac{0}{0+488} = 0$
- F1 - Class 1(Sleepy) as Positive: $\frac{0}{0+\frac{488+0}{2}} = \frac{0}{0+244} = \frac{0}{244} = 0$
- Precision - Class 0(Non Sleepy) as Positive: $\frac{851}{851+488} = \frac{851}{1339} = 0.6355$
- Recall - Class 0(Non Sleepy) as Positive: $\frac{851}{851+0} = \frac{851}{851} = 1$
- F1 - Class 0(Non Sleepy) as Positive: $\frac{851}{851+\frac{0+488}{2}} = \frac{851}{851+244} = \frac{851}{1095} = 0.7772$

- Accuracy: $\frac{0+851}{0+0+488+851} = 0.6355$

2.3 Question 3

Due to the dataset being unbalanced, since there are much more NSL recordings than SL recordings the metric used should be the weighed average metric.

2.4 Question 4

Trained with run_baseline_students.py and svm_functions.py.

Various options and combinations for the parameters were tested. While experimenting with the *class_weight* parameter yielded better results in some cases, since it compensated for the unbalanced dataset. In the end, the best parameter combination didn't alter the *class_weight* values.

2.5 Question 5

In every dataset we tried the following 3 kernels: Linear, RBF, Polynomial. With each kernel we varied values of C, γ and d whenever they would affect the SVM performance with said kernel.

The Linear Kernel performed poorly. Even while altering the value of C (slack) to increase the amount of outliers permitted during the training process, it was not able to produce good results.

On the other hand, the Radial Basis Function and Polynomial kernels performed similarly with the RBF kernel having slightly better results.

In both of them increasing C to have more of a hard margin SVM would always overfit to the training data, while decreasing it would see the results decline. The best value of C for both kernels was found to be around $C = 0.8$.

The γ parameter only influences the RBF kernel. It is common to use $\gamma = \frac{1}{\#features}$. In our case altering it to any different number would diminish results or make the model overfit to the training data.

The d parameter only affects the polynomial kernel being the degree of the kernel. A kernel with small degree would lead to a model not being able to describe the data well. A kernel with a high degree would overfit to the training data. The value of d found to be the best one was $d = 3$.

The AVEC2013 dataset performed slightly better than the eGemaps featureset but didn't outperform the IS11 feature set.

	Hyperparameters				Train				Development			
	Kernel	C	γ	d	Acc	PPV	Recall	F1	Acc	PPV	Recall	F1
IS11	RBF	0.8	$\frac{1}{\#features}$	-	0.919	0.918	0.912	0.915	0.740	0.724	0.694	0.702
eGeMaps	RBF	0.8	$\frac{1}{\#features}$	-	0.869	0.864	0.862	0.863	0.723	0.703	0.709	0.706
AVEC2013	RBF	0.5	$\frac{0.5}{\#features}$	-	0.847	0.841	0.837	0.839	0.732	0.710	0.699	0.703

2.6 Question 6

The files can be found on the folder "predictions" and the model on the main folder

3 Part3

3.1 Question 1

Checked

3.2 Question 2

3.2.1 One Layer

The model has an average performance with the accuracy in both datasets being around 70%.

3.2.2 Adding Layers

With the addition of linear layers the model performs like it only knows the prior distributions of the data. It presents results like the ones in Part2-Question 2.

3.2.3 Adding Batch Normalization and Dropout

The model performance improves dramatically. The model now has better results than the SVM classifier.

3.3 Question 3

3.3.1 eGemaps

- Epochs - 40
- Learning Rate - 0.15

- Weight Decay - 0
- Batch size - 128
- Dropout - 0.1
- Network Structure (#nodes from layer to layer) - $\#features \triangleright 128 \triangleright 64 \triangleright 32 \triangleright \#classes$

3.3.2 IS11

- Epochs - 15
- Learning Rate - 0.15
- Weight Decay - 0
- Batch size - 128
- Dropout - 0.3
- Network Structure (#nodes from layer to layer) - $\#features \triangleright 128 \triangleright 64 \triangleright 32 \triangleright \#classes$

3.3.3 AVEC2013

- Epochs - 11
- Learning Rate - 0.1
- Weight Decay - 0
- Batch size - 128
- Dropout - 0.45
- Network Structure (#nodes from layer to layer) - $\#features \triangleright 128 \triangleright 64 \triangleright 32 \triangleright \#classes$

	Train				Development			
	Acc	Precision	Recall	F1	Acc	Precision	Recall	F1
IS11	0.829	0.823	0.817	0.820	0.756	0.737	0.724	0.730
eGemaps	0.840	0.859	0.810	0.822	0.740	0.719	0.719	0.719
AVEC2013	0.829	0.844	0.801	0.812	0.749	0.731	0.713	0.719

3.4 Question 4

The files can be found on the folder "predictions" and the model on the main folder

4 Additional Feature sets

In addition to the eGemaps and the IS11 feature sets we also experimented with the feature set from the AVEC challenge in 2013. The results on this dataset will be presented in the tables in conjuction with the 2 asked datasets

5 Part4

The files can be found on the folder "predictions"