EC:447 Pattern Recognition and Machine Learning

Vowel-only versus consonant sound classification using EEG data corresponding to speech prompts

Submitted by:

Group K

Aditya Nishtala
Akarsh P
D R V Ramesh
Keerthan Shagrithaya
Sandeep Kotta

Department of Electronics and Communication Engineering
National Institute of Technology Karnataka
Surathkal,Karnataka, India

Contents

1 Features

Our data matrix size is 869 x 11160. There are 347 data points corresponding to vowel-only class and 522 corresponding to consonant class. We have divided the entire data into 80%:20% for training and testing respectively.

We have used three different types of features.

- LDA feature : During the training phase, we reduced the 11160 dimensions to a single dimension using LDA. The same model was used to project points belonging to the validation and test data.
- Features of importance : For the entire train data, we selected 4 features of importance from 11160, and divided by their corrsponding standard deviations. The importance score was calculated as follows

$$I(k) = \frac{|\mu_k^{(1)} - \mu_k^{(2)}|}{|\sigma_k^{(1)} - \sigma_k^{(2)}|} \tag{1}$$

  where,
  $\mu_k^{(i)}$ is the mean of all the values corresponding to the k^{th} feature of points belonging to i^{th} class
  $\sigma_k^{(i)}$ is the standard deviation of all the values corresponding to the k^{th} feature of points belonging to i^{th} class
- Non linear features : We also considered non linear features of the above mentioned features of importance. Upto 5^{th} order features were used. The non linear features were divided by their standard deviations before giving as input to classifier.

2 Learning curves for Logistic regression

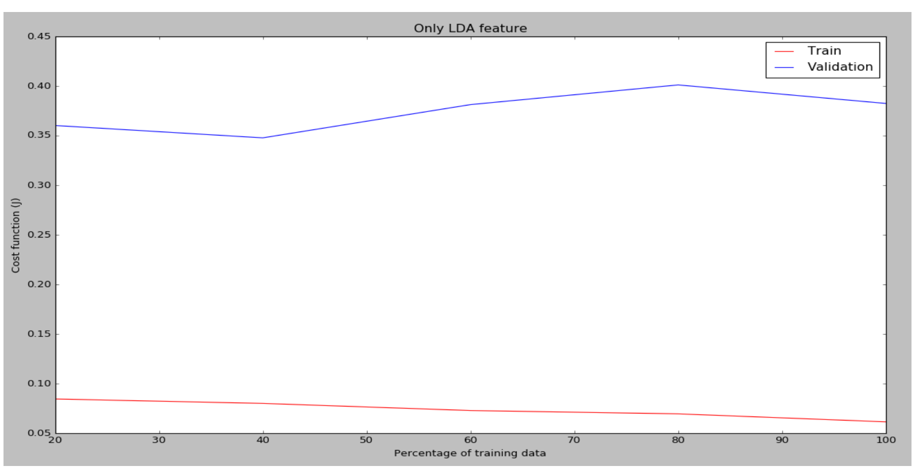

Figure 1: Learning curves when only LDA features were used

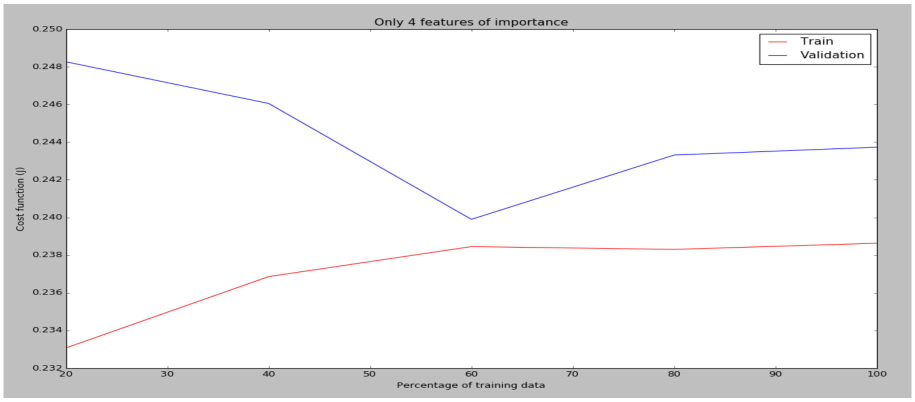

Figure 2: Learning curves when only 4 features of importance were used

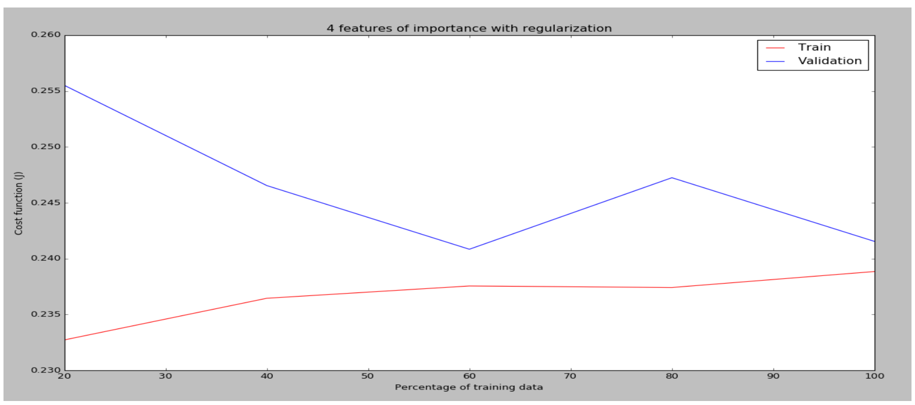

Figure 3: Learning curves when features of importance were used with regularization $\lambda = 1000$

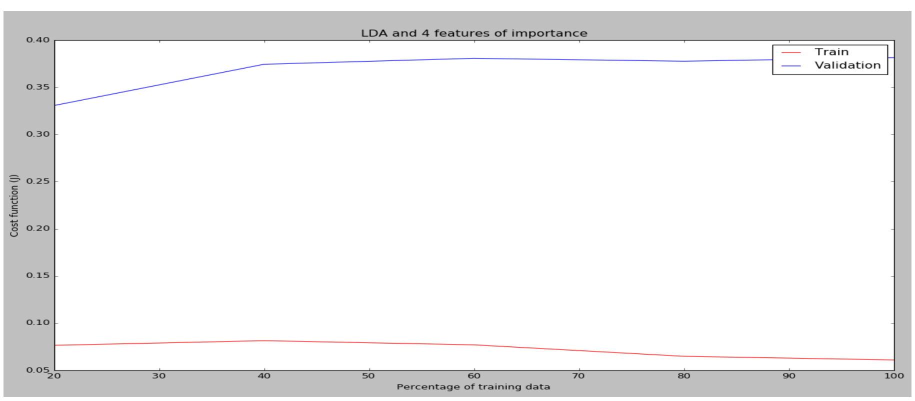

Figure 4: Learning curves when LDA features and 4 features of importance were used

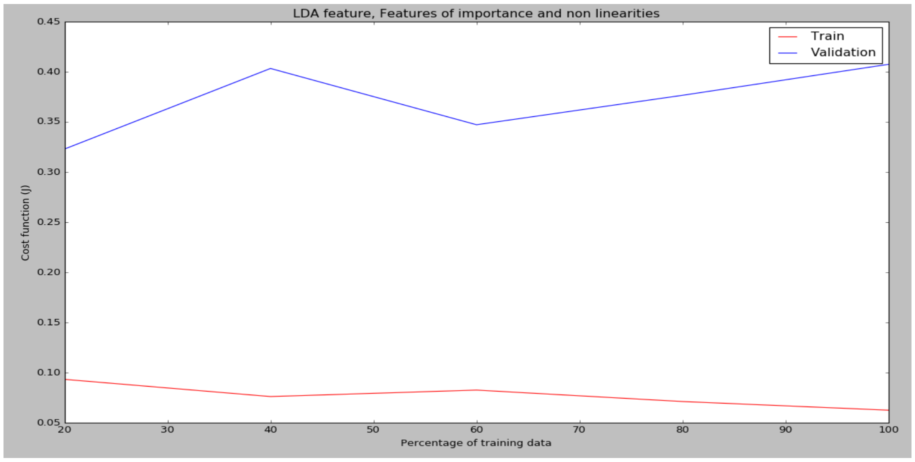

Figure 5: Learning curves when LDA features, 4 features of importance and their non linearities (2 - 5 order) were used

Figure 1 has a very low training error and high validation error. This indicates overfit. The nature of the curves are such that the classifier isn't learning much with more data. Figure 2 exhibits the required learning curve characteristics. The classifier isn't overfit but both train error and validation error are high and almost same. Figure 3 is very similar to Figure 2. We understand that regularization penalty has no effect on the result. Whenever LDA features are used, we find that the train error becomes very low and the validation error is very high compared the train error. This is seen in Figure 4 and Figure 5.

3 Learning curves for GMM

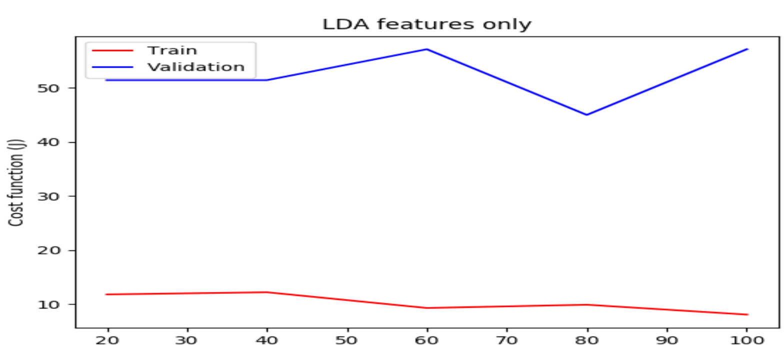

Figure 6: Learning curves when LDA features are used for GMM

Figure 6 shows that it isn't learning with data. Train error is very low compared to validation error.

4 Test data results

Only LDA features	53.75%
Only 4 features of importance	59.537%
4 features of importance with regularization	59.537%
LDA and 4 features of importance	51.45%
LDA, 4 features of importance and their non linearities	57.80%

Table 1: Logistic regression results

LDA features	56.648%

Table 2: GMM classifier results

References

[1] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. Vol. 2. New York: Wiley, 1973.

[2] Walt, Stéfan van der, S. Chris Colbert, and Gael Varoquaux. ”The NumPy array: a structure for efficient numerical computation.” Computing in Science & Engineering 13.2 (2011): 22-30.

[3] Jones, Eric, Travis Oliphant, and Pearu Peterson. ”SciPy: open source scientific tools for Python.” (2014).

[4] Pedregosa, Fabian, et al. ”Scikit-learn: Machine learning in Python.” Journal of Machine Learning Research 12.Oct (2011): 2825-2830.