Language Recognition in Speech

Harsh Manocha, Magus Verma

Motivation

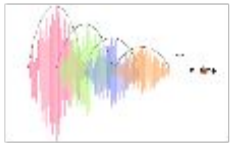
- Problem definition: Recognizing language in speech data
- Huge amount of work has been done in speech recognition (speech to text)
- But requires languages to be known
- A very important step towards seemless translation of audio content across languages
- Can also help in telephone calls
- What if it can be automated?

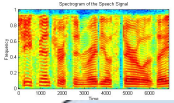
The Approach

Obtaining data through scripts

Preprocessing: Stripping, Smoothening

Dividing into sliding windows





Spectral Centroid,Spectral Entropy, MFCCs, Zero Crossing Rate, Energy

Feature extraction: MFCC, Delta

$$\Delta_{t,i} = \frac{c_{t+1,i} - c_{t-1,i}}{2}$$

$$\Delta^2_{t,i} = \frac{c_{t-1,i} - 2c_{t,i} + c_{t-1,i}}{4},$$

Splitting into Training and Testing

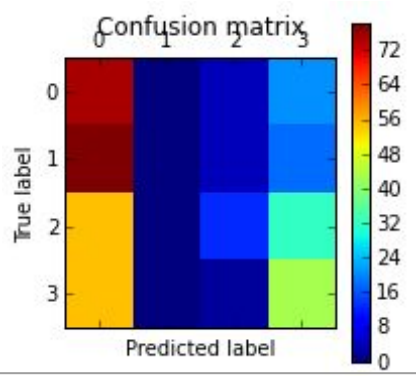
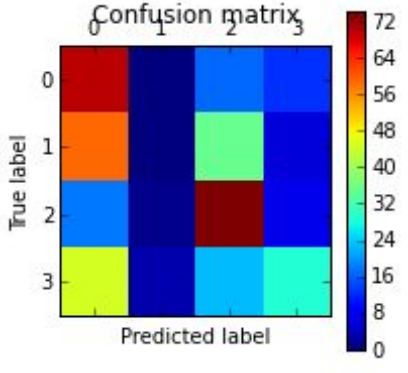
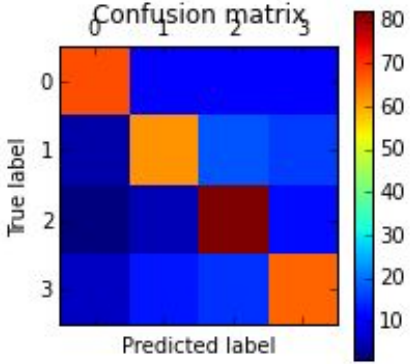
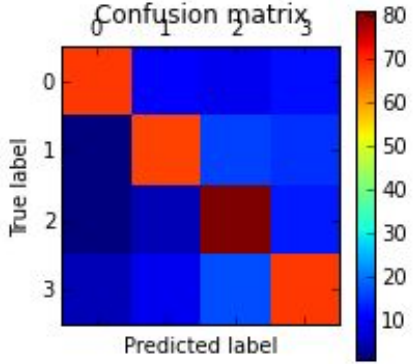
Training and Predicting Classification of samples

Linear SVC

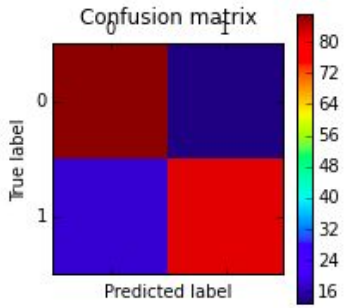
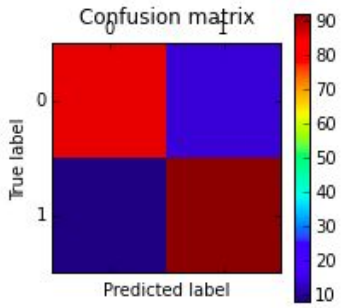
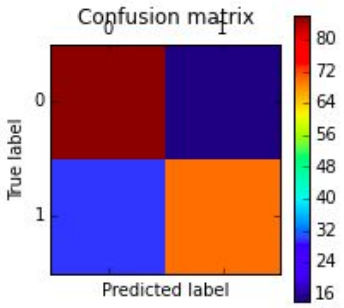
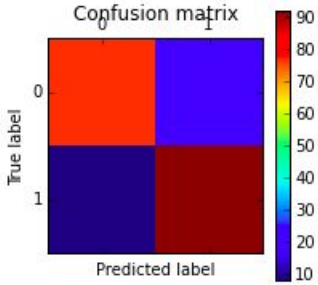
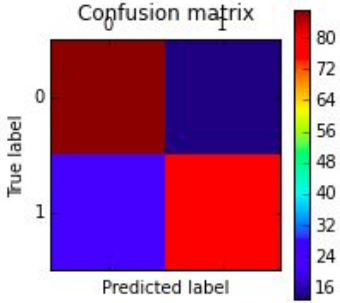
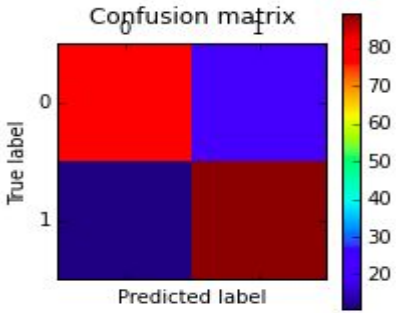
RDF

Detecting language by majority voting

Results (4-class)

Classifier	Accuracy (with Hamming)	Accuracy (without Hamming)
SVM	32.75 % 	43.25 % 
RDF	69.50 % 	71.75 % 

	ENGLISH		FRENCH		GERMAN		SPANISH	
	SVM	RDF	SVM	RDF	SVM	RDF	SVM	RDF
ENGLISH			50%	85%	54.5%	**89%**	54.5%	81%
FRENCH	50%	85%			50.5%	84.5%	62%	**78%**
GERMAN	54.5%	89%	50.5%	84.5%			50%	84.5%
SPANISH	54.5%	81%	62%	78%	50%	84.5%		

 Language 0: GERMAN ; Language 1: SPANISH	 Language 0: GERMAN ; Language 1: ENGLISH	 Language 0: SPANISH ; Language 1: FRENCH
 Language 0: GERMAN ; Language 1: FRENCH	 Language 0: SPANISH ; Language 1: ENGLISH	 Language 0: FRENCH ; Language 1: ENGLISH

Challenges

- Obtaining dataset
  - Needed variety of speakers and different languages
  - Spent majority of time
- Dataset very noisy
  - Too much background noise in audio recording
  - Speakers not speaking naturally (monotonic)
- Diving into the realm of audio world