Training Support Vector Machines for SAE Vowel Phoneme Classification

Alan J. Zaffetti

University of Massachusetts Amherst

azaffett@umass.edu

April 23, 2015

Overview

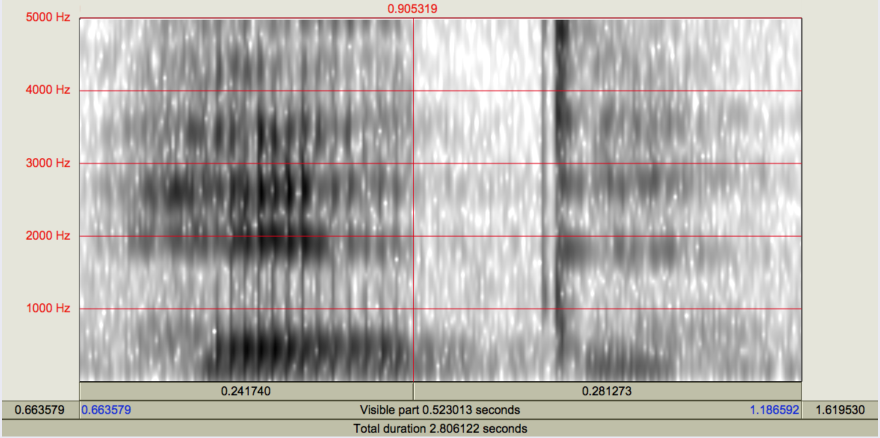
Vowel Phonemes

- Phonemes are the units of sound in an utterance.
- Vowel phonemes are found at nucleus of syllable.
sh**i**p	b**oo**k	**e**gg
c**a**t	c**u**p	h**o**t
- Here are some more of them:

ae="had"	ah="hod"	aw="hawed"	eh="head"
er="heard"	ei="haid"	ih="hid"	iy="heed"
oa="boat"	oo="hood"	uh="hud"	uw="who'd")

Classifying Phonemes

Praat Spectrogram



Classifying Phonemes

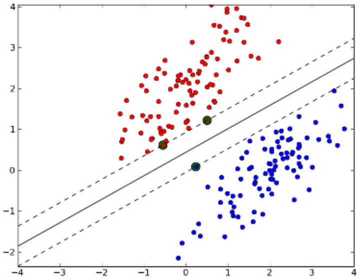
Phoneme Classification

"The task of deciding what the phonetic identity of a speech utterance is." [Boujelbene et. al., 2008]

- In speech recognition this means learning a model from data.
- One such model is Support Vector Machines.

SVM

1. A binary linear classifier capable of learning from example.
2. Represents data as points in space and partitions them over a line (via MMP).



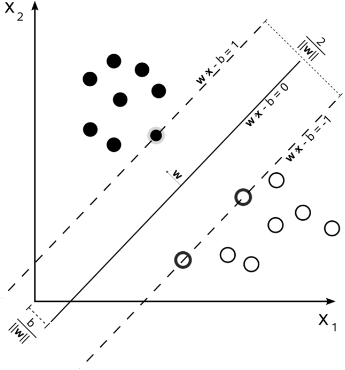
- Linear equations partition space in $\Re^2$.
- **+1** class and -**1** class.

MMD Principle

If the training data are linearly separable:

1. Select two lines in a way that they separate the data and there are no points between them.
2. Try to maximize their distance.

This leads to lesser generalization error. Leaves room for new data, noise.

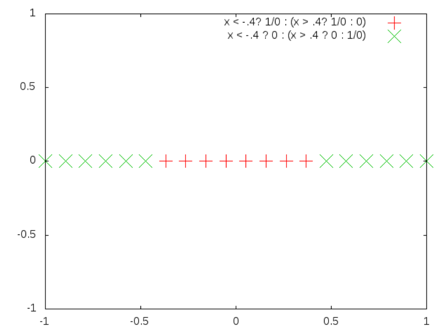


- The graph on the left represents data in $\Re^1$.

  Are the two classes linearly seperable?

  Is it possible to draw a straight line?

Strange Data



Linear Seperability

- Yes! If you cheat.
- Map each data point

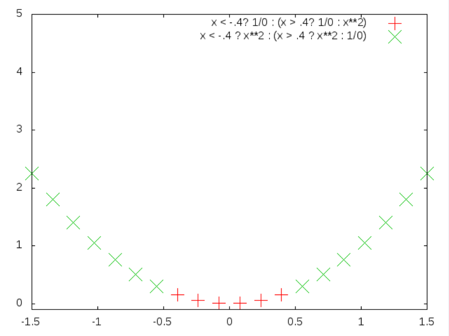
$$x \longrightarrow (x, x^2) \qquad (1)$$

Theorem ($\Re \rightarrow \Re^n$ Mapping)

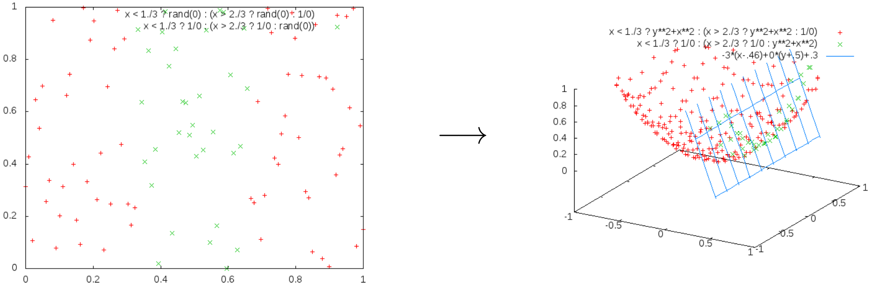
This result generalizes:

$$\forall D \exists n : \exists \text{ a linear cut in } \Re^n \quad (2)$$

Strange Data



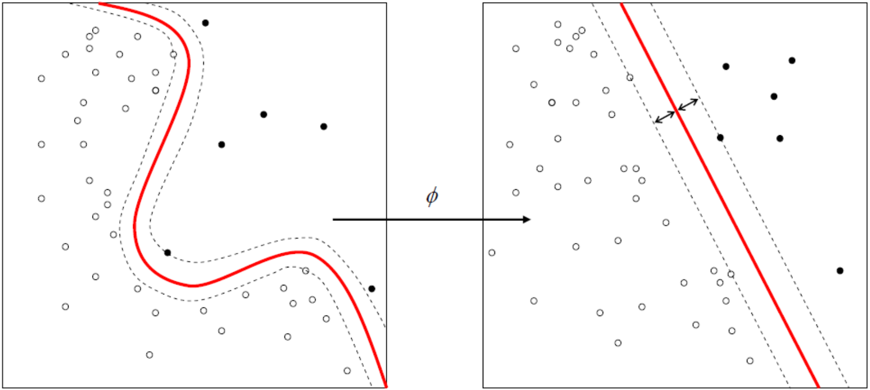
Linear Seperability



Kernel Trick

Kernel Functions

Φ-mapping functions exist which map data to even higher dimensions, and don't share the same limitations as $\Re$-mapping.



Project Goals

- Hillenbrand Vowel Formant dataset
- Contains information on vowel length, formant frequency, speaker data.
- Tune an SVM to recognize major vowel distinctions.

SVM Feature Model

Formally:

$$(\Delta t, f0, \hat{f1}, \hat{f2}, \hat{f3}, f4) \longrightarrow v \in V \tag{3}$$

Where Δt is the length of the vowel, and $f_0 \cdots f_4$ are formant data at different vowel durations, v is a vowel class in V.

Project Challenges

Challenge

SVM are binary classifiers. They must be extended to work with multiple classes.

Possible Solutions

1. Create $O(n^2)$ classifiers for each vowel pair $(i \neq j) \implies (i, j)$ and chose the class which performs best on its $O(n)$ tests.
2. Create $O(n)$ cascading classifiers for each vowel v, testing $(v, \bar{v})$. Select the *best* value.

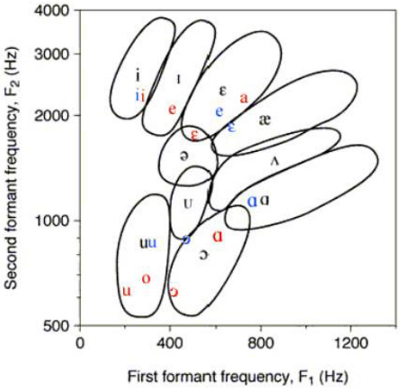
Challenge

Vowel classification is hard.

Project Challenges

Possible Solution

SVMs accel at high-dimensional problems. More dimensionality could make the distinctions more clear, albeit more computationally intensive to decide.



References

Boujelbene et. al. (2008)

Vowel Phoneme Classification Using SMO Algorithm for Training Support Vector Machines

Information and Communication Technologies: From Theory to Applications IEEE, 1 – 5.

The End