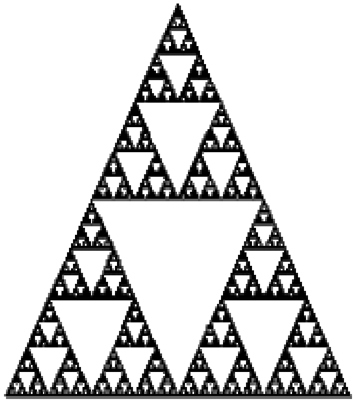
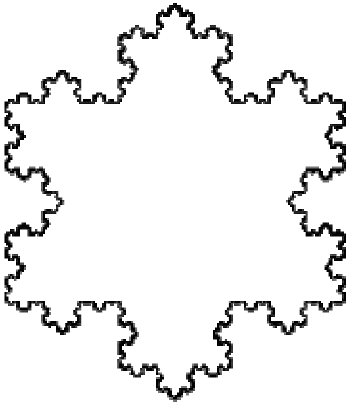
RMIT



Fractal Theory & Naïve Bayes Classifier - A Basic Introduction



Sridhar P. Arjunan

School of Electrical and Computer Engineering,

RMIT University

GPO Box 2476V, Melbourne, VIC 3001,

Australia

s3099587@student.rmit.edu.au

22 Sep. 2006

Outline

- Introduction - Fractals
- Self- Similarity
- Fractal Dimension – As a Feature
- Naïve Bayes Classifier
- An Example

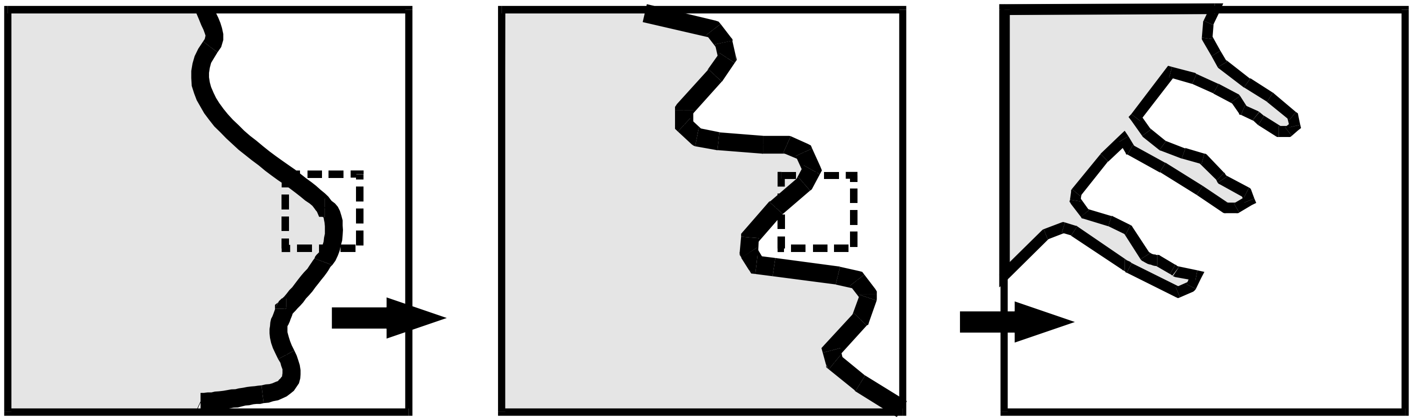
s3099587@student.rmit.edu.au

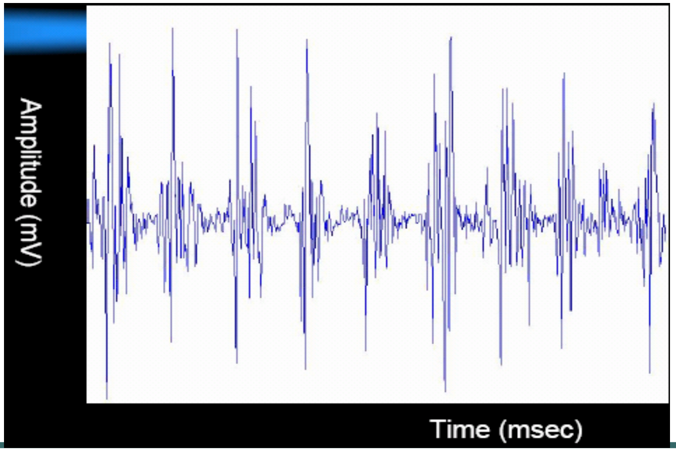
Introduction- Fractals

- Fractals refer to
  - objects or patterns that have fractional dimension.
  - objects exhibit self similarity.
  - objects or patterns on magnification will yield a structure that resembles the larger structure in complexity.
- The measured property of the fractal process is scale dependant and has self similar variations in different time scales and produces a broad band frequency spectrum

s3099587@student.rmit.edu.au

Fractal





s3099587@student.rmit.edu.au

Self-Similarity

Self similarity is a distinctive feature of most fractals;

- a small portion of the figure resembles some larger part when magnified either closely or exactly. It can also be called scale invariance.

- Self-similarity, in a strict sense, means that the statistical properties of a stochastic process do not change for all aggregation levels of the stochastic process.

- That is, the stochastic process *"looks the same"* if one zooms in time *"in and out"* in the process.

s3099587@student.rmit.edu.au

Self-similarity in EMG

- In complex bio signals like EMG, there exists self similarity phenomenon, in which there is a small structure (MU) that resembles the larger structure.
- Fractal dimension can be applied to determine this self similarity.
- Burst within burst behaviour of EMG in time has the property that patterns observed at one sampling rate; say one ms are statistically similar to patterns observed at a slower sampling rate, say one s.
- These nested patterns can be described using the concept of self-similarity, a key property of fractal objects. Exactly self-similar fractal objects are identical regardless of the scale or magnification at which they are viewed.

s3099587@student.rmit.edu.au

Test for self similarity

- Define the *'aggregated process'* of the time series of the EMG signal. The aggregated process is a new time series generated by averaging the original time series over non overlapping blocks of size, say '*m*'.

$$y^{(m)}(k) = \frac{1}{m}\sum_{l=0}^{m-1} y(km - l)$$

- The process is *self similar* if the variance of the aggregated process decays *slowly* with *m* that is,

$$Var(y^{(m)}) \approx m^{-\beta} \qquad \text{with} \qquad 0 < \beta < 1$$

and $H = 1 - \beta / 2$

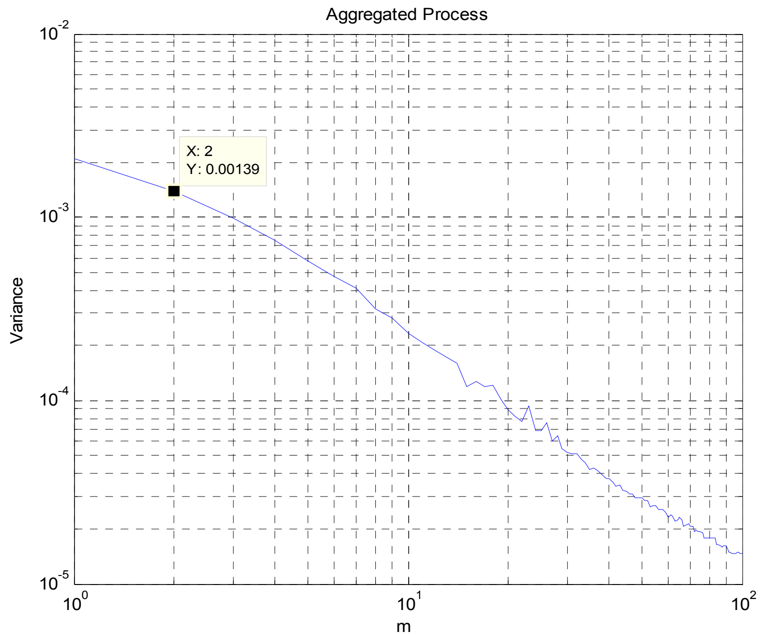
where H expresses the degree of self similarity; large values indicate stronger self-similarity.

s3099587@student.rmit.edu.au

Self similarity analysis of EMG signal

•Self similarity analysis of *EMG signal recorded from wrist flexion* was performed to ensure the self similar structure. The new time series signal was generated as a '*aggregated process*'.

•The logarithmic plot was performed between the variance of the process and the size *'m'.* From the plot, it's seen that the variance is decaying slowly with 'm'.



β = (log(0.002099)-log(0.0002316))/(log(10)-log(1)) ;

β = 0.9573 < 1 and H=0.5213

s3099587@student.rmit.edu.au

Various Approaches

- Three new approaches has been proposed in literature for characterisation SEMG. The methods that characterise the EMG spectral distribution
  - Logarithmic representation of EMG spectrum
  - Poisson representation of EMG spectrum and
  - the method that examines the ‘complexity’ of raw EMG i.e. Fractal dimension of EMG.
- Raw EMG signal has characteristics that are fractal. Because it is self – similarity over a range of scales and the statistical properties of a part are proportional to those of the whole.
- Fractal dimension of EMG has been found sensitive to magnitude and rate of force of generation.

s3099587@student.rmit.edu.au

Fractal Dimension

- The fractal dimension represents degree of self- similarity mathematically. Fractal dimension has several different definitions. In determining fractal dimensions of EMG, we use definition of fractal dimension as follows . Power spectrum of a signal with fractal characteristic has the following form:

$$S(\omega) \propto \frac{1}{\omega^{\alpha}}$$

in which α is a real number.

- For one-dimension time series signal, fractal dimension is defined as:

$$D_f = (5 - \alpha)/2$$

s3099587@student.rmit.edu.au

Fractal Dimension and EMG

- If lower fractal dimension is achieved when the muscles contract, it shows discharge units can act more synchronously and have closer correlation, and the whole muscle system assume more harmony.
- If higher fractal dimension is achieved when the muscles relax, it shows discharge units act asynchronously and the muscles approach the state of entire relaxation more closely in time of no motion.
- Fractal Dimension measures the degree of fractal boundary fragmentation or irregularity over multiple scales (ie transients in EMG signal)
- Fractal dimension may be useful as alternative means to evaluate the EMG and Evoked Potential signals. In bio-medical, waveforms showing repetitive patterns (ECG, EEG, EMG) are often analyzed in the terms of Fractal Dimension.

s3099587@student.rmit.edu.au

EMG and Wavelets

- EMG signals are non stationary multi component signals made of the superimposition of motor unit action potentials (MUAPs) characterised by individual time and frequency localisations.
- Moreover, under certain conditions, the various MUAPs can be considered the scaled versions of a single wave. MUAPs due to deeper motor units are dilated with respect those due to more superficial ones.
- Therefore the global signal can be modelled as the superimposition of delayed and scaled versions of a basic component.
- Wavelet transform represents a very suitable analysis method for this class of signals. This fractal dimension can be determined using Wavelets because of its duality property with fractals.

s3099587@student.rmit.edu.au

General view

Important Characteristics of Features

- Discrimination
  - How good are the features
- Reliability
  - How reliable is the decision rule
- Independence
  - Features should be uncorrelated with each other
- Small number
  - Complexity in recognition increases with the number of features used

Classification

- View the recognition problem as that of generating *"decision boundaries"* separating *m classes* on the basis of the observed vector

s3099587@student.rmit.edu.au

Naïve Bayes Classifier

- Naïve Bayes Classifier technique is based on so- called Bayesian Theorem and particularly suited when dimensionality of inputs is high.
- Naïve Bayes Classifiers can handle an arbitrary number of independent variables whether continuous or categorical.
- Naïve Bayes Model is a simple and well known method for performing supervised learning of a classification problem.
- Naïve Bayes can be modelled in several different ways including normal, lognormal, gamma and Poisson density functions

s3099587@student.rmit.edu.au

Naïve Bayes Model

- If we consider Y to be an object to be classified, then Bayes' Theorem can be read as a formula for the probability that Y belongs in category Xi.

$$P[X_i | Y] = \frac{P[Y|X_i] \cdot P[X_i]}{\sum_j P[Y|X_j] \cdot P[X_j]}$$

- Assuming that the conditional probabilities for different Xi differ, it is reasonable to simply assign Y to the Xi having the highest conditional probability (i.e., conditional upon the attribute values of Y).
- Since the denominator in Bayes' Theorem is independent of i (and is always nonnegative), the numerator of the most likely Xi will also have the greatest magnitude.
- Thus, to perform classification, we need only compute the numerator in Bayes' Theorem for each Xi and then pick the Xi giving the largest value.

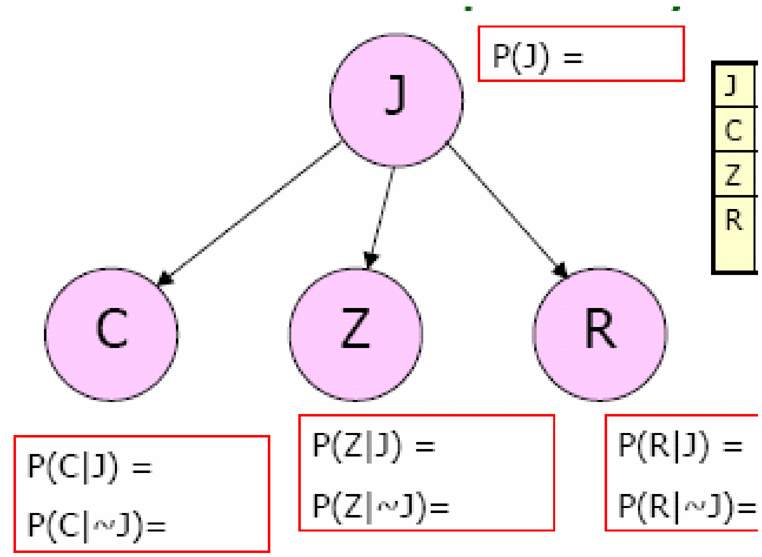
s3099587@student.rmit.edu.au

Contd..

- The problem has now been reduced to computing

$$P[Y|Xi]*P[Xi]$$

- P[Xi] can be trivially estimated by counting the training examples that fall into Xi and dividing by the size of the training set.
- P[Y|Xi] is less trivial, and it is the computation of this term that warrants use of the word naive in the phrase Naive Bayes Model.
- The model is *naive* in the sense that it assumes (often unjustifiably) that the attribute values of object Y are independent.



s3099587@student.rmit.edu.au

How NBC works????

- The NB classifier selects the most likely classification V_{nb} given the attribute values $a_1, a_2, \ldots\ldots, a_n$

  this results in

$$V_{nb} = \arg\max_{v_j \in V} P(v_j) \prod P(a_i \mid v_j)$$

$$where$$

$$P(a_i \mid v_j) = \frac{n_c + mp}{n + m}$$

where

n = the number of training examples for which $v = v_j$

n_c = number of examples for which $v = v_j$ *and* $a = a_i$

p = a priori estimate for $P(a_i|v_j)$

m = the equivalent sample size

s3099587@student.rmit.edu.au

A General Example

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Attributes are *colour, type, origin* and the subject, stolen can either be *yes* or *no*.

s3099587@student.rmit.edu.au

Training Example

- For example , if we have to classify the class Red Domestic SUV????

- We need to calculate the probabilities

$$P(Red|Yes),\ P(SUV|Yes),\ P(Domestic|Yes)\ ,$$

$$P(Red|No)\ ,\ P(SUV|No),\ \text{and}\ P(Domestic|No)$$

s3099587@student.rmit.edu.au

Contd...

$$P(Red|Yes) = \frac{3 + 3 * .5}{5 + 3} = .56 \qquad P(Red|No) = \frac{2 + 3 * .5}{5 + 3} = .43$$

$$P(SUV|Yes) = \frac{1 + 3 * .5}{5 + 3} = .31 \qquad P(SUV|No) = \frac{3 + 3 * .5}{5 + 3} = .56$$

$$P(Domestic|Yes) = \frac{2 + 3 * .5}{5 + 3} = .43 \qquad P(Domestic|No) = \frac{3 + 3 * .5}{5 + 3} = .56$$

We have $P(Yes) = .5$ and $P(No) = .5$, so we can apply equation (2). For $v = Yes$, we have

```
P(Yes) * P(Red | Yes) * P(SUV | Yes) * P(Domestic|Yes)

    =  .5 * .56 * .31 * .43   = .037
```

and for $v = No$, we have

```
P(No) * P(Red | No) * P(SUV | No) * P (Domestic | No)

    = .5 * .43 * .56 * .56 = .069
```

Since $0.069 > 0.037$, our example gets classified as 'NO'

s3099587@student.rmit.edu.au

Conclusion

- The Naive Bayes Model is clearly an easy approach to supervised learning of classification tasks.
- On test problems, one will often find that it performs less well than other methods, such as back propagation neural networks, but occasionally, it will perform better than one or more competing techniques.
- The method performs best when attribute values approach independence.
- For problems where attributes have many complex interactions, there is less reason for optimism.
- However, the Naive Bayes Model is a good candidate for a first attempt at learning a new classification task.

s3099587@student.rmit.edu.au

THANK YOU………..

s3099587@student.rmit.edu.au