Bayesian Data Modelling For Smarter Applications

Paul Walmsley paul@pjwhams.co.uk
London OSJam 18 June 2009

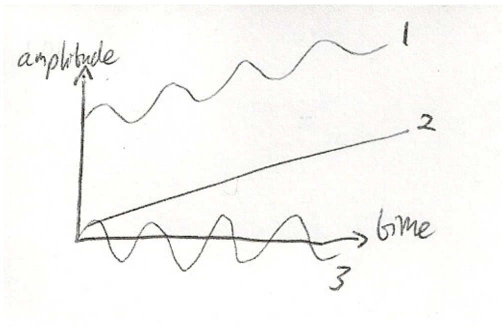
Motivation

- One avenue towards creating great, compelling software is to build in intelligence
- One of the ways of building in intelligence is by use of pattern recognition, eg gesture recognition, smile detection in cameras, adapting to the user's behaviour
- This is a golden age for sensors: touch, motion, tilt, cameras, microphones

Data Modelling and the General Linear Model

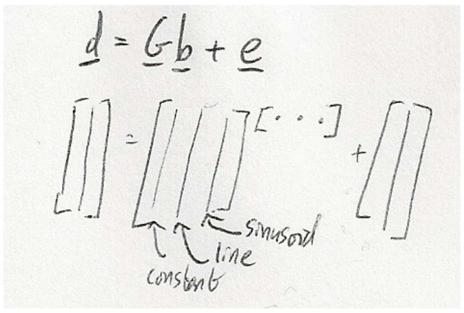
Traditionally, ('non-parametric') signal processing techniques have been about filtering and transforming signals to extract the 'features' of interest.

For example, if we have a signal (1) comprised of a ramp (2) with a sinusoidal variation (3) then we might pass the signal through a lowpass filter to isolate the ramp and a filterbank to isolate the sinusoid.



A *data modelling* ('parametric') approach is by contrast constructional, where you build a model to represent the data and then find the parameters that give the best fit to the data.

The General Linear Model is a versatile way of representing models in vector form. **d** is a vector of the observed data, **G** is a matrix composed of a number of columns where each contains a type of geometric function ('basis functions') that you want to



detect, $\mathbf{b}$ is a vector containing the amount of each of the basis functions, and $\mathbf{e}$ is an error term.

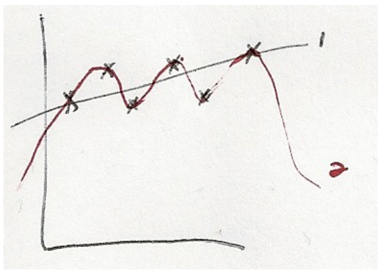
Usually you want to minimise the error term (i.e., find the model parameters that give the best fit to the data), and you can solve this to give the *least squares* solution, $\hat{\mathbf{b}}$

$$\hat{\underline{b}} = (\underline{G}^T\underline{G})^{-1}\underline{G}^T\underline{d}$$

There are two interesting special cases: if the basis functions are a constant and a straight line then the least squares solution gives you the linear regression coefficients. If the basis functions are sinusoids the least squares solution gives you the FFT.

The power in this model comes from the ability to represent much more complex basis functions that can vary over time and which may have other unknown parameters, eg time-varying frequencies or multiple changepoints.

Bayesian modelling



Do we always want to find the solution that gives the smallest error? If we have six noisy data points (right) we can identify a linear trend and draw a straight line through the middle. Or we could create a more complicated curve that goes through all points. That would minimise the error but gives us a model that is too complicated for the small amount of data we've seen. The least squares solution often *overfits* the data.

This is where Bayes theory comes in. The least squares solution is the one that maximises the *likelihood* $p(\mathbf{d}\,|\,\boldsymbol{\theta})$ (θ represents all parameters) but the better solution is the one that maximises the *posterior* probability as this takes account of the *prior* probability.

$$\underbrace{p(\underline{\theta}|\underline{d})}_{\text{posterior}} = \frac{\overbrace{p(\underline{d}|\underline{\theta})}^{\text{likelihood}} \cdot \overbrace{p(\underline{\theta})}^{\text{prior}}}{p(\underline{d})}$$

The prior has two functions:

- Penalises complex models, so a simpler model would be picked unless a lot of data justified it

- Represents all your prior information, so you can represent knowledge about the problem domain (e.g., user tends to shake the sensor at a rate of about 6Hz), or adapt to the user's previous behaviour (e.g., user draws gesture A more often than gesture B)

Bringing the General Linear Model and Bayes Together

We can combine the descriptive abilities of the GLM with Bayesian priors to create an incredibly powerful data model. *Unfortunately*, that makes it too complex to solve algebraically. *Fortunately* there are some techniques such as Markov Chain Monte Carlo where we can infer the model parameters by using a random number generator to explore the model space. This is computationally expensive, but may be worthwhile if you need a very sophisticated model.

Conclusion

- Data modelling is a powerful concept that has great potential for smart, adaptive, interactive applications
- Bayesian priors can also be used to harness real-world experience about the problem domain and to adapt to the user's past behaviour
- The resulting system can be lightweight for speed or heavyweight for complex data analysis

For more information, including some papers and my thesis, which used these techniques for polyphonic pitch estimation in audio see http://www.pjwhams.co.uk/research.html