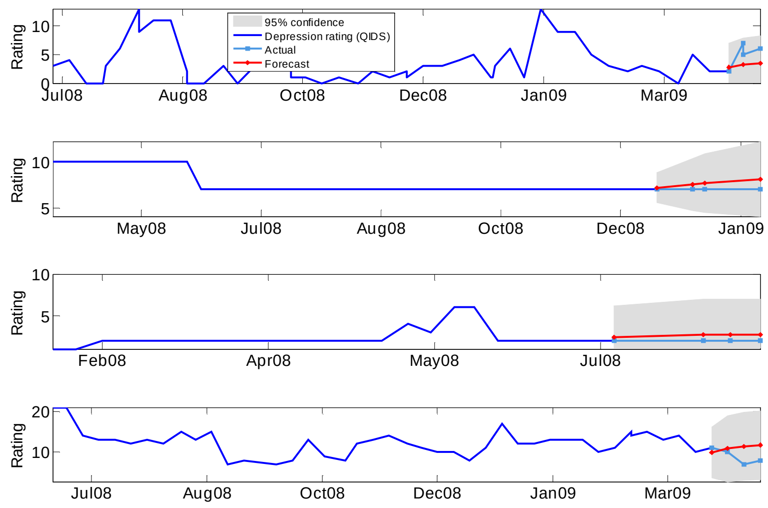
Predicting mood using Gaussian Process Regression

[1]PJ Moore, [1]PE McSharry, [1]MA Little , [2]JR Geddes, [2]WM Stevens

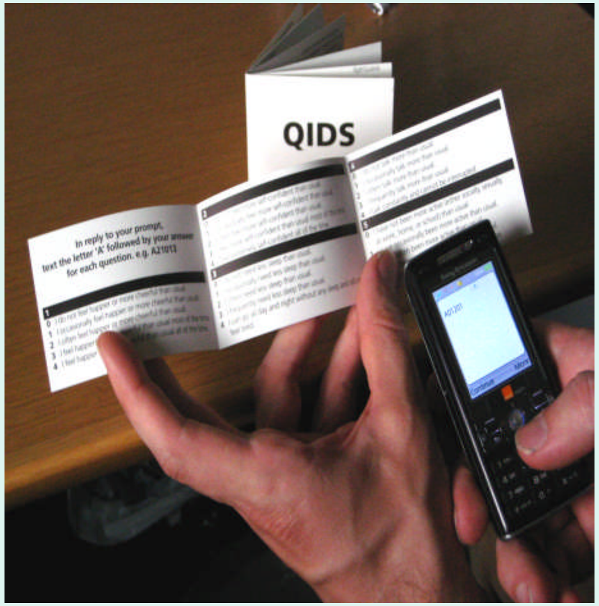
[1]Mathematical Institute, University of Oxford. [2]Department of Psychiatry, University of Oxford.

1. Introduction

The Department of Psychiatry at Oxford University has developed an automated mood monitoring system for patients with bipolar disorder. The system is based on text message and web-based mood rating scales to allow remote data capture. Currently, about 200 patients use the system to record their mood rating score each week. Using anonymised data from these patients, we are developing mathematical techniques and algorithms for the forecasting of mood events such as the onset of depression.



Four examples of mood estimation



2. Methods

We model a mood time series $\boldsymbol{y}$ $\{y_1..y_n\}$ as a function $f(\boldsymbol{x})$ which relates time indices $\{x_1..x_n\}$ to random variables $\{Y_1..Y_n\}$ representing mood. For a *Gaussian Process* the joint distribution between any finite number of these variables is Gaussian. The process is specified by a mean function and a covariance function:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$$
$$k(\mathbf{x},\mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$$

We choose the mean as zero and use a parameterised covariance function such as a rational quadratic:

$$k_{RQ}(x_p, x_q) = \left(1 + \frac{(x_p - x_q)^2}{2\beta l^2}\right)^{-\beta}$$

where x_p, x_q are time indices, and β,l are free *hyperparameters*.

Values of the hyperparameters are chosen by maximising the likelihood $p(\boldsymbol{y} \mid \boldsymbol{x}, \beta, l)$

The joint distribution between known ratings $\boldsymbol{y}$ and unknown test ratings $\boldsymbol{f}_*$ with time indices $\boldsymbol{x}_*$ is given by

$$p\begin{bmatrix}\mathbf{y}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K + \sigma_n^2 I & K_*^T \\ K_* & K_{**}\end{bmatrix}\right)$$

Where K, K_{**} are the covariance matrices for $\boldsymbol{y}$ and $\boldsymbol{f}_*$ respectively and K_* is the covariance matrix for $\boldsymbol{y}$ with $\boldsymbol{f}_*$. σ^2 is a noise term. The conditional distribution $p(\boldsymbol{f}_* \mid \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{x}_*)$ is then

$$\mathcal{N}(K_*(K + \sigma_n^2 I)^{-1}\mathbf{f}, K_{**} - K_*(K + \sigma^2 I)^{-1} K_*^T)$$

and the predictive equations are

$$\bar{\mathbf{f}}_*(\mathbf{x},\mathbf{x}_*) = K_*(K + \sigma_n^2 I)^{-1}\mathbf{y}$$
$$\mathbb{V}((\mathbf{f}_*(\mathbf{x},\mathbf{x}_*)) = K_{**} - K_*(K + \sigma_n^2 I)^{-1} K_*^T$$

In the figure (above left) the mean function is visualised in red, with a 90% confidence interval in grey.

3. Results

Next Step Forecast method	RMSE % Min	RMSE % Mean	RMSE % Max
Autoregressive	3.3	14.3	41.1
Exponential Smoothing	1.1	13.1	35.2
Gaussian Process	4.0	13.1	32.9

We measure performance by predicting the next mood rating for different training sets and measuring the root mean square error. Results are compared with two other forecasting approaches (above). Future work will use multivariate forecasting, automatic determination of input relevance, and k-step prediction.

Key Words: Gaussian process, time series analysis.

Oxford Centre for Industrial and Applied Mathematics (OCIAM), Mathematical Institute, University of Oxford, UK Email: paul.moore@maths.ox.ac.uk URL: http://people.maths.ox.ac.uk/moorep