PRML 本の 6 章に詳細あり。式番号は、PRML 本に合わせている。

1 Gaussian process

とりあえず難しい数式の話は後回しね。データポイント$x_1, \cdots, x_N$に対して、真の値を$\boldsymbol{y} = y_1, \cdots, y_N$とすると、各値は以下の確率で与えられるとする。

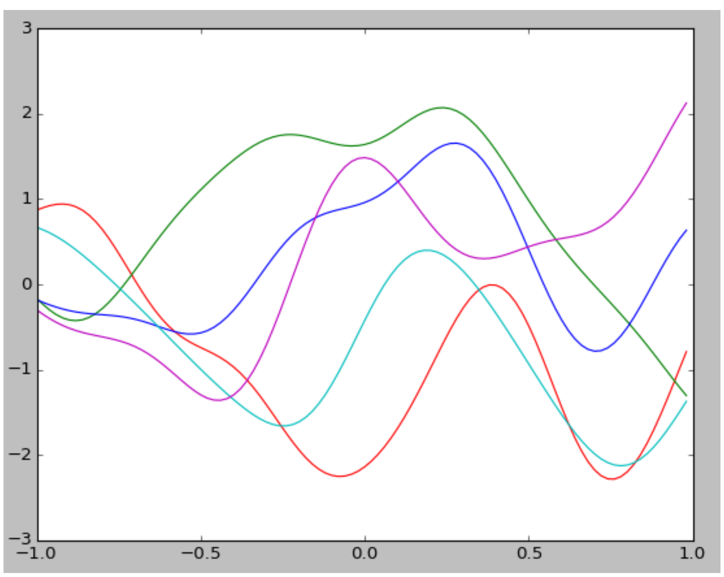
$$p\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}\right) \sim N\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, Cov\right)$$

ただし、

$$Cov = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix}$$

※なんで、平均が 0 なの？ってところがすごい混乱するんだけど、そういうものと考えるしかないみたい。逆に言うと、元の値が偏っていたりする場合は、平均が 0 になるよう調整する必要がある。

下図に、いくつかの曲線が表示されている。１つの曲線が、１つの$\boldsymbol{y}$に相当する。ガウス分布に従った乱数なので、平均 0 から少し離れたりしているわけだよね。



ガウス過程を使った regression では、与えられた観測値に対して、最も近い曲線を当てはめるというイメージだ。※ハッキリ言って、Gaussian process そのものは良く分からんけど、次章の regression さえ理解しておけば問題ないと思う。

2 Gaussian process for regression

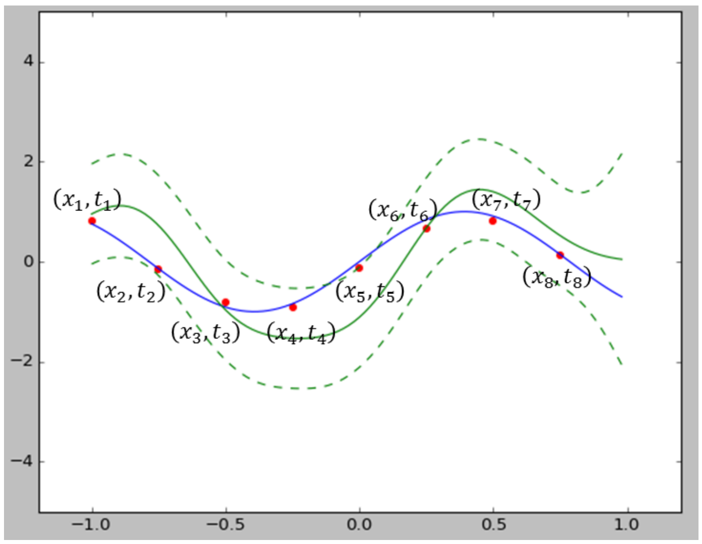
とりあえず難しい数式の話は後回しにする。データポイント$x_1, \cdots, x_N$に対して、真の値を$y_1, \cdots, y_N$、観測値$\boldsymbol{t} = t_1, \cdots, t_N$とする。観測値は、真の値に対して分散$\beta^{-1}$のノイズが加味されると考える。つまり、

$$p(t_n|y_n) = N(t_n|y_n, \beta^{-1}) \tag{6.58}$$

また、データポイントの共分散を以下のように定義する。

$$k(x_n, x_m) = \theta_0 \exp\left\{-\frac{\theta_1}{2}\|x_n - x_m\|^2\right\} + \theta_2 + \theta_3 x_n^T x_m \tag{6.63}$$

なんでこんな式を共分散として使用するのか？よく使われる式みたいなので、とりあえず使っとけって感じ。



この時、任意の点x_{N+1}の値$\hat{y}_{N+1}$は、次のようにして推定される。

$$\hat{y}_{N+1} = \boldsymbol{k}^T C_N^{-1} \boldsymbol{t} \tag{6.66}$$

ただし、

$$k = \begin{bmatrix} k(x_1, x_{N+1}) \\ k(x_2, x_{N+1}) \\ k(x_3, x_{N+1}) \\ \vdots \\ k(x_N, x_{N+1}) \end{bmatrix}$$

また、

$$C_N = \begin{bmatrix} k(x_1, x_1) + \beta^{-1} & k(x_1, x_2) & \cdots & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) + \beta^{-1} & \cdots & k(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & \cdots & k(x_N, x_N) + \beta^{-1} \end{bmatrix}$$

※式(6.66)は、こう考えると理解できる。点$\boldsymbol{x}_{N+1}$の値は、既に観測したデータ群$\boldsymbol{t}$の加重平均的な感じで計算され、その重みは、点$\boldsymbol{x}_{N+1}$に近い点の観測値をより大きい重みにする。すごく自然な考え方だよね。

推定された値の精度は、以下のように偏差を計算して推定できる。

$$\sigma^2 = c - k^T C_N^{-1} k \tag{6.67}$$

ただし、

$$c = k(x_{N+1}, x_{N+1}) + \beta^{-1}$$

とりあえず、以上の式を使うことで、regression は行えるのだ。とりあえず、「使う」分にはこれで困らないよね。