RERGRESSIONS

Polynomial Regression

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + ... + w_M x^M = \sum_{j=0}^{K} w_j x^j$$

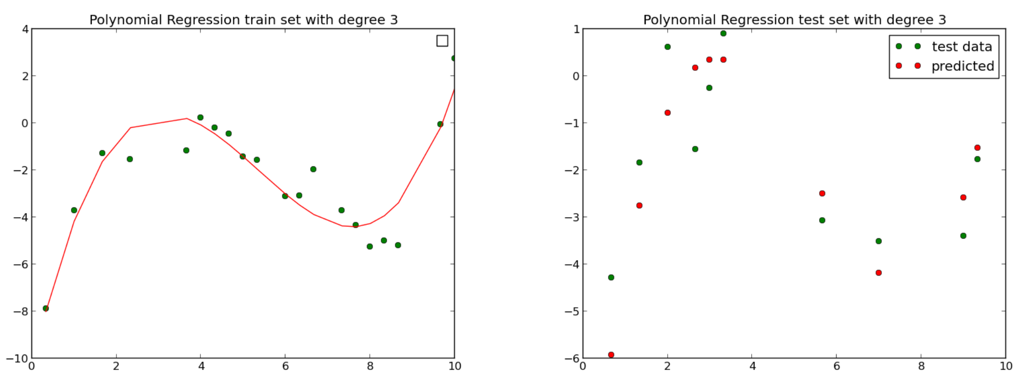
where $y(x, \mathbf{w})$ is predicted output, x is data, $\mathbf{w}$ is weight, and K is degree of the polynomial.

The weight can be found in a closed form:

$$w = \varphi(X)^{-1} t$$

where $\varphi(x)_i j = x_j^i$, i ith degree for jth data.

For the practice purpose, I used toydata that I sampled from the degree 3 polynomial with gausian noise added. Sampled 60 points, and selected random 40 points as train data and random 20 points as test data. After fitting degree 3 curve into the train data.



The Left hand side of image is curve fitted on train data, and right image is the data points predicted from test data.

Bayesian Curve Fitting using Evidence Approximation

Given $X = (x_1, ..., x_n)^T$ and $t = (t_1, ..., t_n)^T$ as train data (x on x axis, t on y axis), the probaility fitting curve into the data is

$$\begin{aligned} p(t|w, \mathbf{w}, \beta) &= N(t|y(x, \mathbf{w}, \beta^{-1})) \\ &= \prod_{n=1} N(t|y(x_n, \mathbf{w}), \beta^{-1}) \end{aligned}$$

Also, the prior of weight can be presented

$$\begin{aligned} p(\mathbf{w}|\alpha) &= N(\mathbf{w}|0, \alpha^{-1}\mathbf{I}) \\ &= (\frac{\alpha}{2\pi})^{(M+1)/2} exp(-\frac{\alpha}{2}\mathbf{w^T w}) \end{aligned}$$

where α is the precision of the distribution, its a hyperparameter. Then we can predict the test points by

$$\begin{aligned} p(t|x, \mathbf{x}, \mathbf{t}) &= \int p(t|x, \mathbf{w}) p(\mathbf{w}|\mathbf{x}, \mathbf{t}) d\mathbf{w} \\ &= N(t|m(x), s^2(x)) \end{aligned}$$

where $m(x)$ is the mean, $s^2(x)$ is the variance. Because both likelihood and prior function are gaussian, $p(t|x, \mathbf{x}, \mathbf{t})$ becomes gaussian.

$$\begin{aligned} m(x) &= \beta\varphi(x)^T \mathbf{s} \sum_{n=1}^{N} \varphi(x_n) t_n \\ s^2(x) &= \beta^{-1} + \varphi(x)^T \mathbf{S} \varphi(x) \\ \mathbf{S}^{-1} &= \alpha\mathbf{I} + \beta \sum_{n=1}^{N} \varphi(x_n)\varphi(x_n)^T \end{aligned}$$

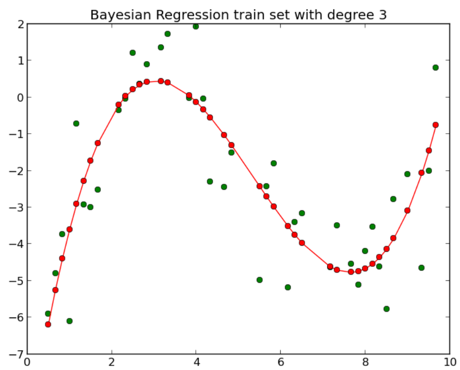
Now, we can use evidence approximation to compute the β, α. The effective parameter γ :

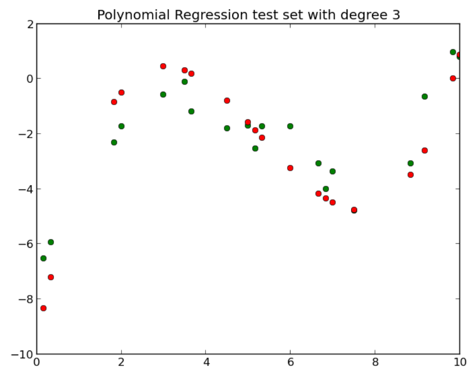
$$\gamma = \sum_i \frac{\lambda_i}{\lambda_i + \alpha}$$

where λ_i is the eigen value of $\mathbf{S}^{-1}$. Then α and β can be computed by

$$\begin{aligned} \alpha &= \frac{\gamma}{m^T m} \\ \beta^{-1} &= \frac{1}{N - \gamma} \sum^{n=1} (t - M_N^T \varphi(x_n))^2 \end{aligned}$$

We can iteratively do this until hyperparameters converges.





The Left hand side of image is curve fitted on train data, and right image is the data points predicted from test data.

Referenced from Christopher M Bishop, Pattern recognition and machine learning