Module-1

Antara Paul, Labani Roy

1 Activity 1: simple regression problem

a. As the polynomial order is increased, the fitting tends to get better. But with further increase, the data points are overfitted. Examples are shown for polynomial degrees of 1, 3 and 9, with 3 being a good fit and 9 being an overfit. For all case, fitting is the best for 1000 data points and worst for 10. Thus, increase in data points helps in fitting. Figs 1-3 show our observations for random x between 0 to 1, and N = 10, 100, 1000.

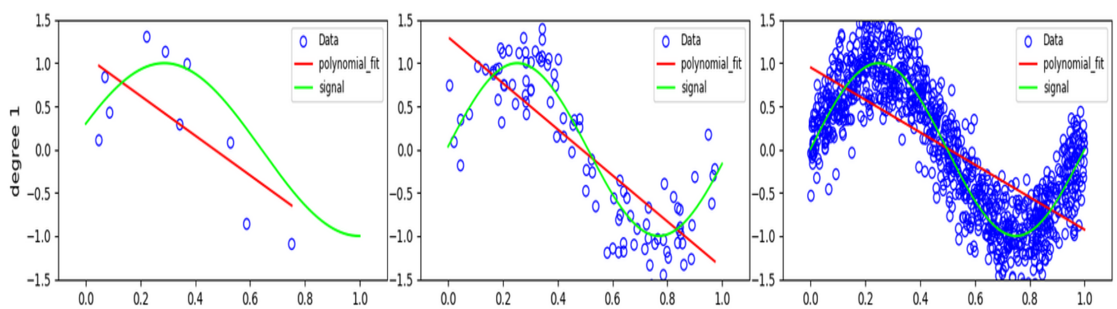

Figure 1

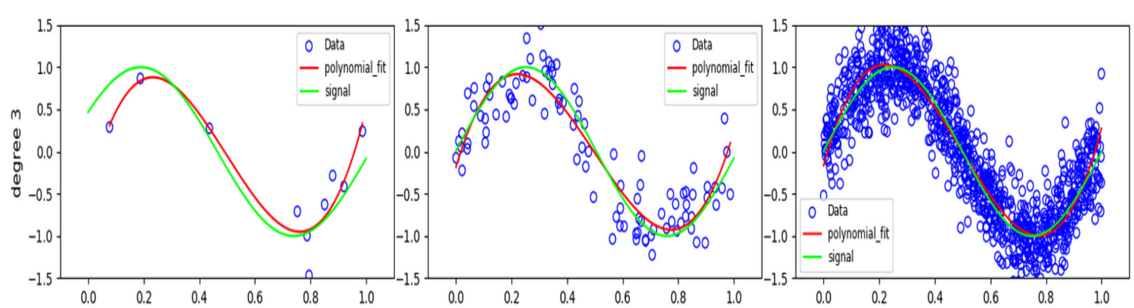

Figure 2

b. For 9th order polynomial with 100 data points, $E_{RMS} = 0.28$ which is close to the noise $\sigma = 0.3$ of our model.

c. We have plotted the E_{RMS} for the training and test data sets. As expected, the test set has a higher error as order of polynomial increases due to over-fitting. Whereas that of the training set decreases, but does not go to zero unlike the textbook because we are fitting for N=100 data points and have only 10 coefficients max.

2 Activity 2: Linear regression with Regularization

a) Linear regression with regularization helps to overcome the overfitting problem by introducing a penalty term λ in the error function. The new error function becomes:

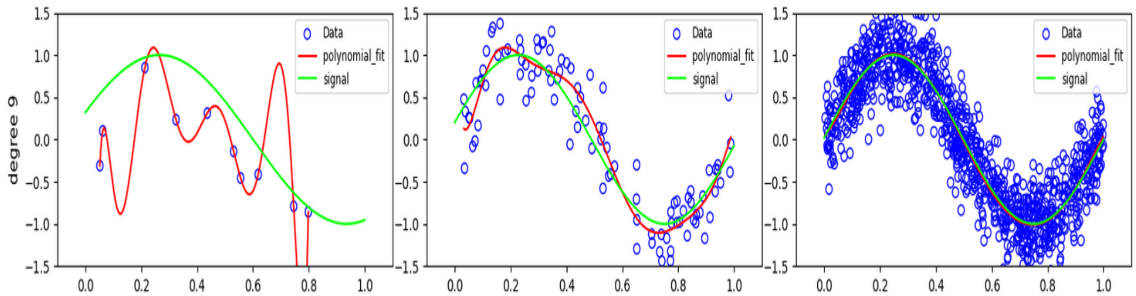

Figure 3

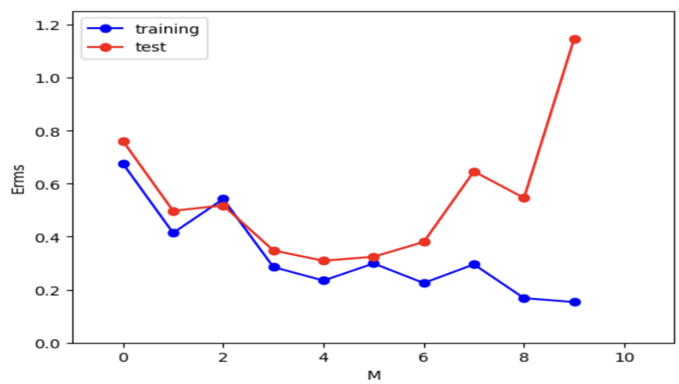

Figure 4

$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\left(y(x_n, \mathbf{w}) - t_n\right)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$.
By minimizing this error function we get the best fitted parameters $\mathbf{w}*$.
If $\lambda = 0$, then $E(\mathbf{w})$ is our normal linear regression error function which shows overfitting problem for higher order polynomial with small sample of data points. But if we increase the λ values, the wild oscillation (1st figure in fig 3) decreases and givs us the following plots 5.

b) Our E_{RMS} plot in Fig. 6 here shows an unexpected trend. The E_{RMS} curves for both sets do increase towards the end but we are not sure why the trends at lower λ are as we got.

3 Activity 3: Practice on Bayes Theorem

Given information are the following , probability of contracting disease A, $P(A) = 0.001$ Now the universal set, $S = A \cup A^c \Rightarrow S_n : n = 1, 2$
$\Rightarrow P(A^c) = 1 - P(A) = 0.999$

probability of getting positive test result given the disease is actually A, $P(+|A) = 0.98$ and probability of getting positive test result given the disease is not A, $P(+|A^c) = 0.03$

Goal is to find the probability of beign contracted with A given the test result is positive, $P(A|+)$.

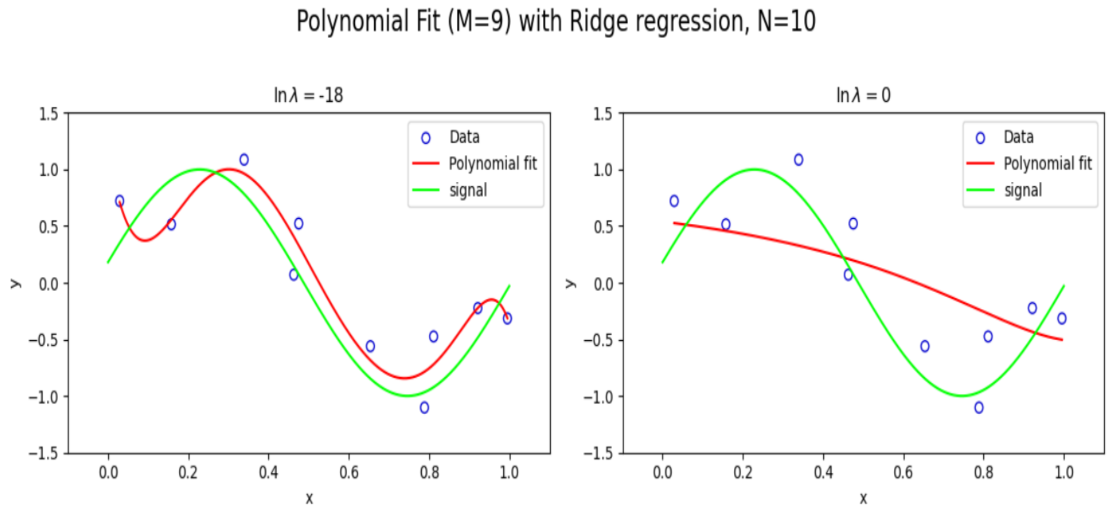

Figure 5

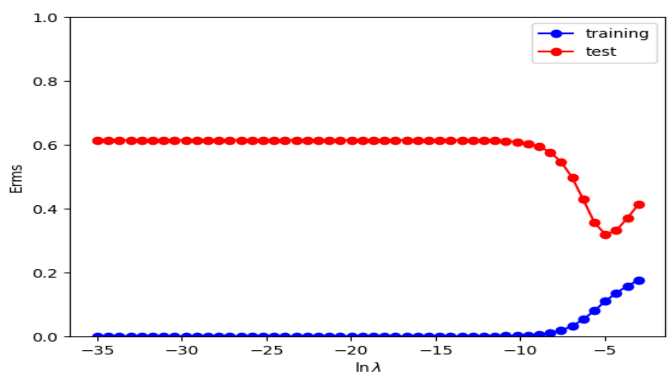

Figure 6

According to the Bayes Theorem,

$$
\begin{aligned}
P(A|+) &= \frac{P(A \cap +)}{P(+)} \\
&= \frac{P(+|A)P(A)}{P(+|A)P(A) + P(+|A^c)P(A^c)}
\end{aligned}
\tag{1}
$$

In the above equation 1, **a)** The likelihood function is $P(+|A)$ and **b)** the prior is $P(A)$. **c)**$P(+)$ is the evidence or the normalization term $\Rightarrow$ $P(+) = \sum_n P(+ \cap S_n) = \sum_n P(+|S_n)P(S_n) = P(+|A)P(A) + P(+|A^c)P(A^c)$

Hence, the value of posterior probability is $P(A|+) = 0.03$. $\Rightarrow P(A^c|+) = 1 - P(A|+) > P(A|+)$.

So, one need not worry about contracting the disease A. :)

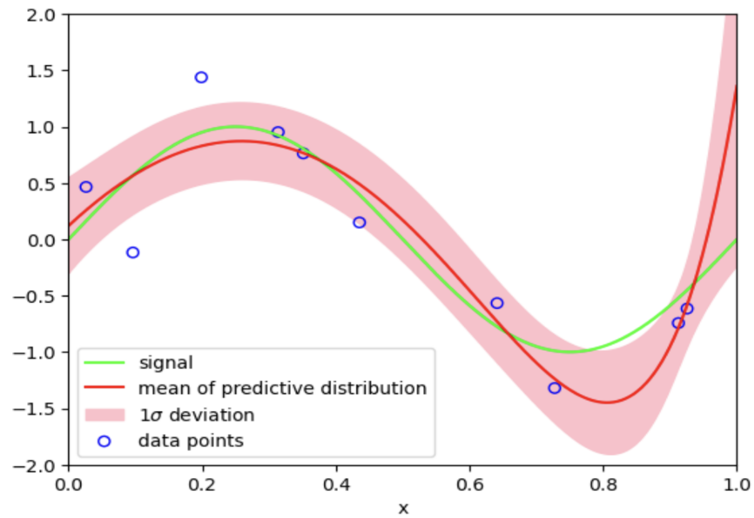

Figure 7

4 Activity 4: Bayesian Curve Fitting

a) We have taken noise with $\sigma = 0.3$ for our sinusoidal signal. β is given as the reciprocal of variance $1/\sigma^2$ and is called precision. This represents the uncertainty that stems from the noise level of the data. For e.g., a higher value of β would imply higher precision, as its reciprocal (i.e, σ^2) would be lower and thus have lesser noise.

b) Comparison between Bayesian approach and frequentist fitting approach

As we can see from Figs 7 and 3, the fitting of 10 data points using the frequentist approach results in over-fitting. However, taking into consideration the mean for the predictive distribution in Bayesian curve fitting gives a good curve-fit.

Moreover, the fitted coefficients are fixed, whereas for bayesian curve fit we get a distribution of fitted parameters (region around 1σ shown in pink here).