(The following images have been taken from Pattern recognition and machine learning- Bishop)

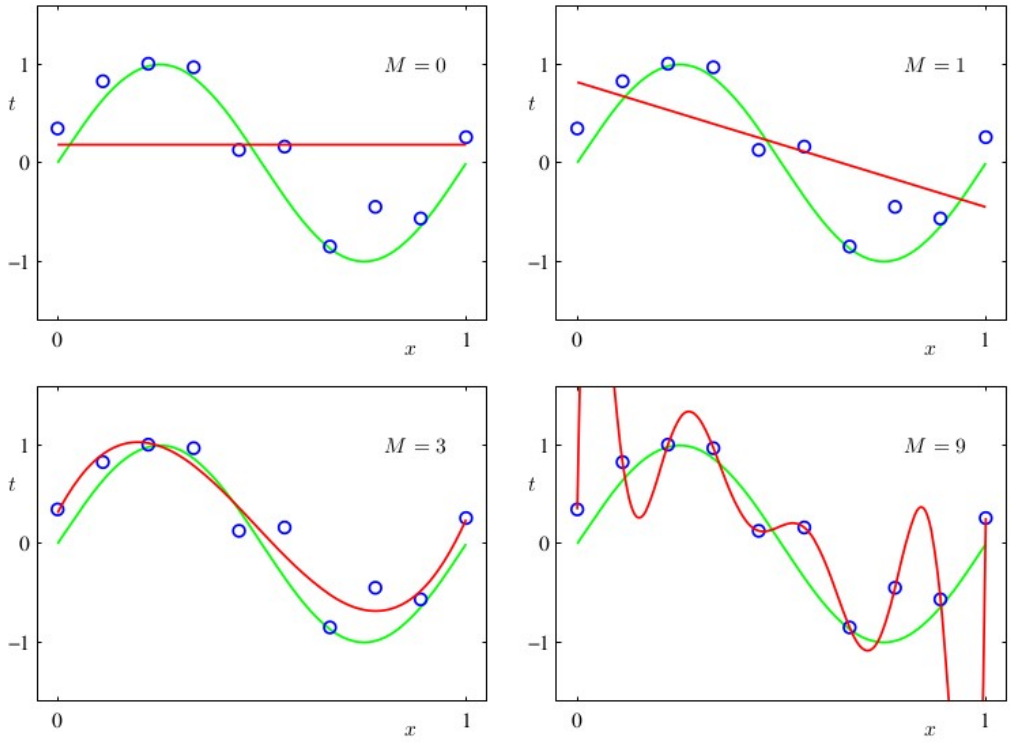

Fig. 1: Plots of polynomials having various orders M , shown as red curves, fitted to the data set.

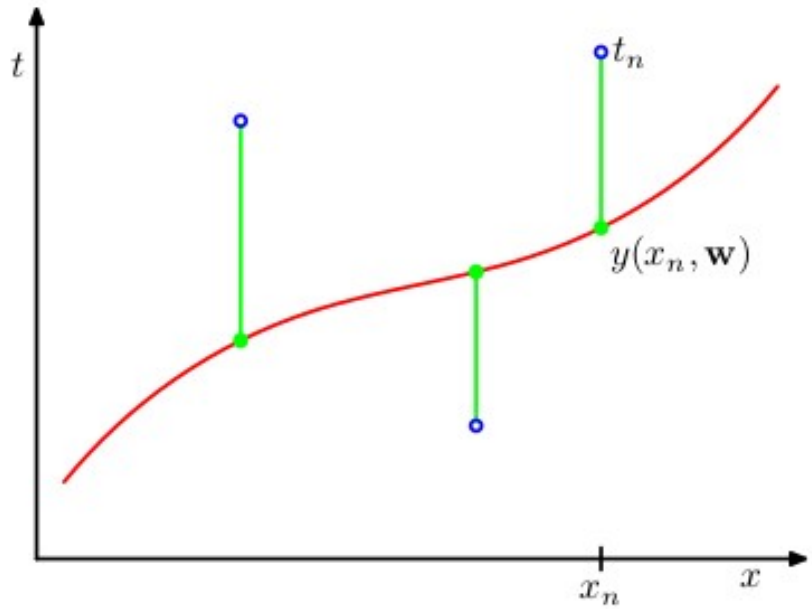

Fig. 2: The error function corresponds to (one half of) the sum of the squares of the displacements (shown by the vertical green bars) of each data point from the function y(x, w).

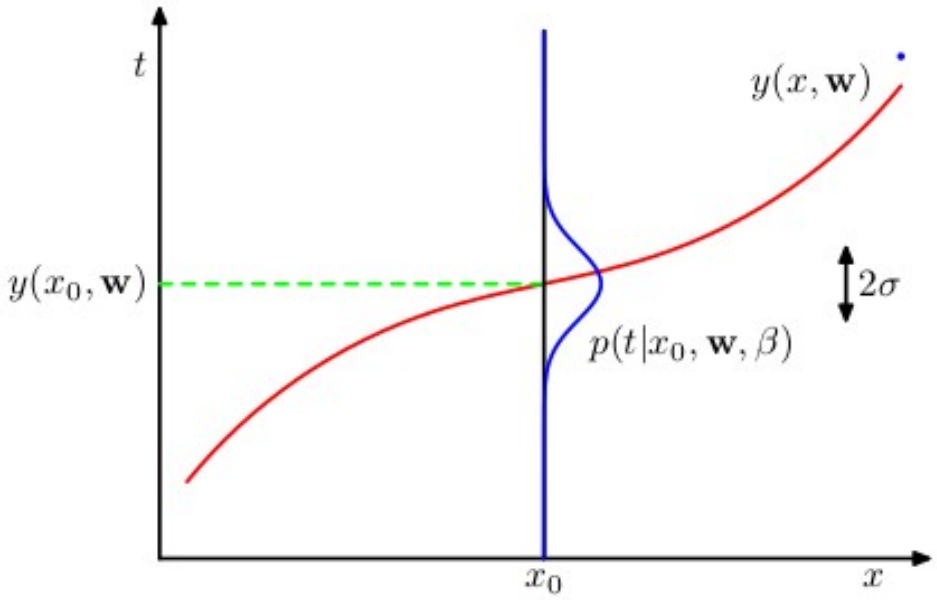

Fig. 3: Schematic illustration of a Gaussian conditional distribution for t given x, in which the mean is given by the polynomial function y(x, w), and the precision is given by the parameter β, which is related to the variance by $\beta^{-1} = \sigma^2$.

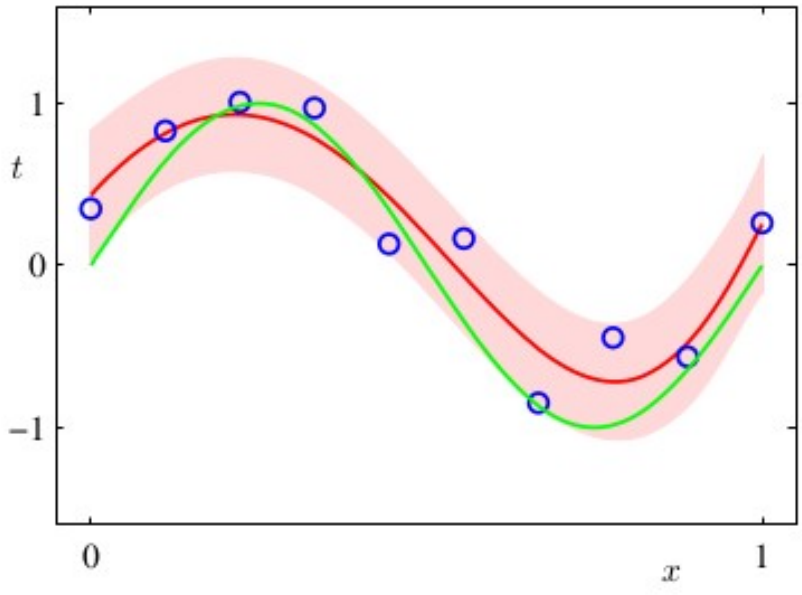

Fig. 4: The predictive distribution resulting from a Bayesian treatment of polynomial curve fitting using an M = 9 polynomial, with the fixed parameters $\alpha = 5 \times 10^{-3}$ and β = 11.1 (corresponding to the known noise variance), in which the red curve denotes the mean of the predictive distribution and the red region corresponds to ±1 standard deviation around the mean.