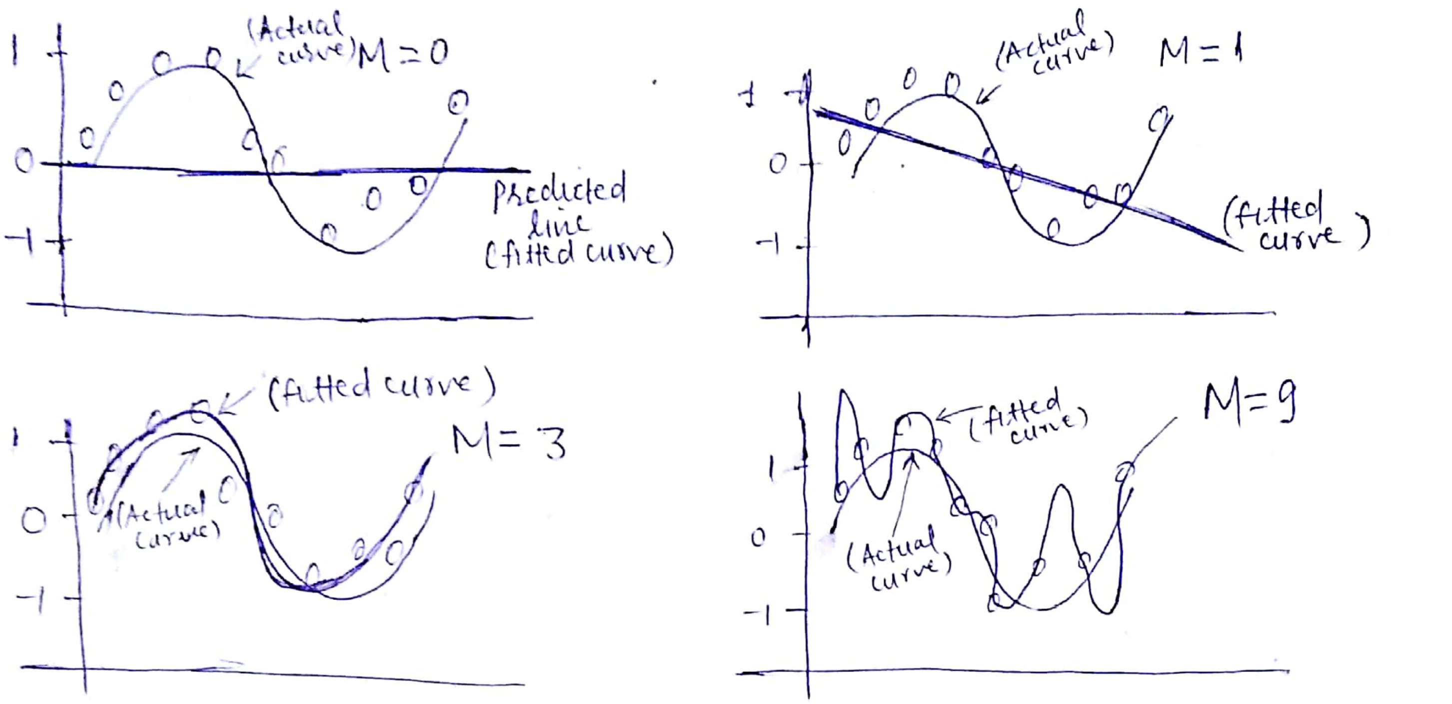
How will we choose M the order of the polynomial? We'll see this will be an important concept called Model selection. In fig we show show four examples of results having orders M = 0, 1, 3 and 9 to the data set.



We notice that M = 0 and M = 1 give rather poor fits to the data. The third (M = 3) polynomial seems to give the best fit to the data. (M = 9) we obtain an excellent fit, in fact curve passes through each data point. Hower the fitted curve oscillates wildly thus gives a very poor representation. This latter behaviour is known as Overfitting.

The goal is to achieve generalization (good) and for this insight we evaluate a separate test set comprising of 100 data points but generated by the same procedure as train set with new different random noise. It is better to evaluate test set performace using root-mean-square error (RMS) defined by:

$$E_{RMS} = \sqrt{2E(w^*)/N} \quad , \quad E(w^*) \text{ for test data.}$$

Here N allows us to compare different sizes of data on equal footing, the square root ensures that E_{RMS} is measured on the same scale (and in the same unit) as the target variable t...

To be continued -

Note - remaining part of this section is in Page 6.