

1 Intorduction Problem Setting

The introduction use polynomial curve fitting as the example.

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j$$

we can use the general defined loss function to describe the fitting performance and the normilized RMS loss function(in order to make compare different dataset size).

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2$$

$$E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^*)/N}$$

We can use the penalty term to control the model complexity in order to match the problem complexity. (As mentioned in class, the DL model can be overparameterized model, which does not follow the traditional model complexity control theory)

$$\tilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

2 Probability Theory

This section is about the genearl probabity theory in discret condition and continus condition.

3 dfsd