1 Intorduction Problem Setting

The introduction use polynomial curve fitting as the example.

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

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).

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

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

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)

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

2 Probability Theory

2.1 Discrete conditoin

2.2 Continuas conditoin

3 The curve fitting revisted

This section firstly introduced general probability theory in discrete conditions and continue conditions. Then with Bayes theorem and gaussian distribution.

4

[1]

5 dfsd

[1]limited to the page limit, you can get teh full version note at https://github.com/Xiang-Pan/NYU_Baysian_Machine_Learning