Lecture 24: Deep neural networks

Professor Ilias Bilionis

Training regression networks - Loss function minimization

The square error loss function for regression

$$x_{1:n} = (x_1, \ldots, x_n) \; ; \; y_{1:n} = (y_1, \ldots, y_n)$$

$$y = f(x; \underline{\theta}) \; : \; DNN \; model$$

parameters

Loss function (Square Error)

error

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left(\underline{y_i} - \underline{f(x_i; \theta)} \right)^2$$

observation   prediction

$$\theta^* = \text{argmin} \; L(\theta)$$

Non-linear optimization method.
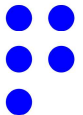
From Gaussian likelihood to loss function

Likelihood: $\quad p(y_i \mid x_i, \vartheta, \sigma) = N\left(y_i \mid f(x_i; \vartheta), \sigma^2\right)$

$$p(y_{1:n} \mid x_{1:n}, \vartheta, \sigma) = \prod_{i=1}^{n} p(y_i \mid x_1, \vartheta, \sigma) = N\left(y_{1:n} \mid f(x_{1:n}; \vartheta), \sigma^2 I_n\right)$$

Max like: $\quad \max_{\vartheta} \log p(y_{1:n} \mid x_{1:n}, \vartheta, \sigma)$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - f(x_i; \vartheta)\right)^2 + \ldots$$

$$\Downarrow$$

$$\min_{\vartheta} L(\vartheta)$$