

2019.11.15.

为什么梯度下降更新参数有效.

$$\Delta w_{hj} = -\eta \frac{\partial E_k}{\partial w_{hj}}$$

$$w_{hj} = w_{hj} + \Delta w_{hj}$$

$$= w_{hj} - \eta \frac{\partial E_k}{\partial w_{hj}} \quad \eta: \text{学习率}$$

$$\frac{\partial E_k}{\partial w_{hj}}: \text{梯度.}$$

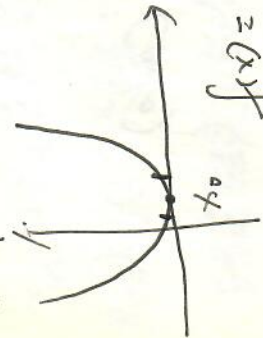
如果一个函数 n 阶可导, 那么我们可以用多项式构造一个相似的函数, 这就是泰勒展开式.

$$\begin{aligned} f(x)_{\text{Taylor}} &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} \times (x-a)^n \\ &= f(a) + \frac{f'(a)}{1!} (x-a) + \frac{f''(a)}{2!} (x-a)^2 + \dots \\ &\quad + \frac{f^{(n)}(a)}{n!} (x-a)^n + R_n(x) \end{aligned}$$

$$f(x)_{\text{Taylor}} = \sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x-x_0)^i, \quad x \in (x_0 - \delta, x_0 + \delta)$$

无限逼近

泰勒展开式 (S Blue Brown) 函数在某点的各阶导数与构造的多项式的对应导数



$y = f(x)$
 $y_0 = f(x_0)$ 相同.

$$f(x) = f(x_0 \pm \Delta x) = y_0 \pm f'(x_0 \pm \Delta x) \Delta x \rightarrow 0$$

$$\text{由于 } y_0 = f(x_0) (x-x_0)^0$$

$$f(x) = f^{(1)}(x_0) (x-x_0)^1 + f'(x_0) (x-x_0)^1$$

$$\text{问题是: } f(x) = \left(\sum_{i=0}^n \frac{f^{(i)}(x_0)}{i!} (x-x_0)^i \right)$$

$$f'(x) = f'(x_0)$$

$$f''(x) = 2 f''(x_0) ??$$

对 $f^{(m)}(x_0) (x-x_0)^m$ 求导看:

$$m! f^{(m)}(x_0) \quad m+1 \text{ 阶导} = 0.$$

—— m 阶导 ——: m 阶导以后都是 0, m 阶导之前 $(x-x_0)$ 都是 0.

$$f(x) = m! f^{(m)}(x_0) + m! f^{(m)}(x_0) (x-x_0)^1 + \dots$$