

深度学习数学基础之高等数学讲义

1. 函数与极限

1.1 函数

函数的定义：

设数集 $D \subset \mathbb{R}$, 则称映射 $f: D \rightarrow \mathbb{R}$ 为定义在 D 上的函数, 通常简记为

$$y = f(x), x \in D,$$

其中 x 称为自变量, y 称为因变量, D 称为定义域, 记作 D_f , 即 $D_f = D$.

1.2. 函数极限的定义

$$y = f(x), \text{ 自变量 } x \text{ 无限趋近的几种形式: } \begin{cases} x \rightarrow x_0 & x \rightarrow \infty \\ x \rightarrow x_0^+ & x \rightarrow +\infty \\ x \rightarrow x_0^- & x \rightarrow -\infty \end{cases}$$

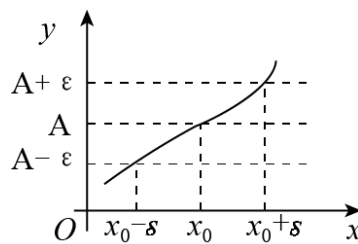
函数极限通常写为: $\lim_{x \rightarrow x_0} f(x) = A$ 或 $f(x) \rightarrow A$ (当 $x \rightarrow x_0$)

函数极限的定义:

如果 $\forall \varepsilon > 0$, $\exists \delta > 0$, 当 $0 < |x - x_0| < \delta$ 时, 总有 $|f(x) - A| < \varepsilon$ 则 $\lim_{x \rightarrow x_0} f(x) = A$

等价于 $\exists \delta > 0$, 当 $x \in \dot{U}(x_0, \delta)$ 时总有 $f(x) \in U(A, \varepsilon)$ 则 $\lim_{x \rightarrow x_0} f(x) = A$.

$x \in \dot{U}(x_0, \delta)$ 的含义: 点的 δ 邻域 $U(\alpha, \delta) = \{x \mid |x - \alpha| < \delta\}$



只要 x 与目标点 x_0 的距离不超过 δ , 那么函数 $f(x)$ 与 A 的距离就不超过 ε

1.3 无穷小与无穷大

无穷小的定义:

如果函数 $f(x)$ 当 $x \rightarrow x_0$ (或 $x \rightarrow \infty$) 时的极限为零, 那么称函数 $f(x)$ 为当

$x \rightarrow x_0$ (或 $x \rightarrow \infty$) 时的无穷小.

特别地, 以零为极限的数列 $\{x_n\}$ 称为 $n \rightarrow \infty$ 时的无穷小.

无穷大的定义:

设函数 $f(x)$ 在 x_0 的某一去心邻域内有定义 (或 $|x|$ 大于某一正数时有定义). 如果对于任意给定的正数 M (不论它多么大), 总存在正数 δ (或正数 X), 只要 x 满足不等式 $0 < |x - x_0| < \delta$ (或 $|x| > X$), 对应的函数值 $f(x)$ 总满足不等式 $|f(x)| > M$, 那么称函数 $f(x)$ 是当 $x \rightarrow x_0$ (或 $x \rightarrow \infty$) 时的无穷大.

1.4 极限的四则运算

定理 1: 两个无穷小的和是无穷小

定理 2: 有界函数与无穷小的乘积是无穷小

定理 3: 如果 $\lim f(x) = A, \lim g(x) = B$, 那么

$$(1) \lim [f(x) \pm g(x)] = \lim f(x) \pm \lim g(x) = A \pm B;$$

$$(2) \lim [f(x) \cdot g(x)] = \lim f(x) \cdot \lim g(x) = A \cdot B;$$

$$(3) \text{ 若又有 } B \neq 0, \text{ 则 } \lim \frac{f(x)}{g(x)} = \frac{\lim f(x)}{\lim g(x)} = \frac{A}{B};$$

定理 4: 设有数列 $\{x_n\}$ 和 $\{y_n\}$. 如果 $\lim_{n \rightarrow \infty} x_n = A, \lim_{n \rightarrow \infty} y_n = B$, 那么

$$(1) \lim_{n \rightarrow \infty} (x_n \pm y_n) = A \pm B;$$

$$(2) \lim_{n \rightarrow \infty} (x_n \cdot y_n) = A \cdot B;$$

$$(3) \text{ 当 } y_n \neq 0 \ (n=1, 2, \dots) \text{ 且 } B \neq 0 \text{ 时, } \lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \frac{A}{B}.$$

定理 5: 如果 $\varphi(x) \geq \psi(x)$, 而 $\lim \varphi(x) = A, \lim \psi(x) = B$, 那么 $A \geq B$.

定理 6: (复合函数的极限运算法则) 设函数 $y = f[g(x)]$ 是由函数 $u = g(x)$ 与函数 $y = f(u)$ 复合而成, $f[g(x)]$ 在点 x_0 的某去心邻域内有定义, 若 $\lim_{x \rightarrow x_0} g(x) = u_0, \lim_{u \rightarrow u_0} f(u) = A$, 且存在 $\delta_0 > 0$, 当 $x \in \overset{\circ}{U}(x_0, \delta_0)$ 时, 有 $g(x) \neq u_0$, 则

$$\lim_{x \rightarrow x_0} f[g(x)] = \lim_{u \rightarrow u_0} f(u) = A.$$

1.5 函数极限求解例题

$$(1) \lim_{x \rightarrow 1} \frac{x^2 - 1}{x - 1} = \lim_{x \rightarrow 1} (x + 1) = 2$$

$$(2) \lim_{x \rightarrow \infty} \frac{1}{x} = 0$$

$$(3) \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$$

题 (3) 解:

$$\because S_{\Delta AOB} \leq S_{\text{扇形} AOB} \leq S_{\Delta AOD}$$

$$\therefore \frac{1}{2} \sin x \leq \frac{1}{2} x \leq \frac{1}{2} \tan x \left(x \in \left(0, \frac{\pi}{2} \right) \right)$$

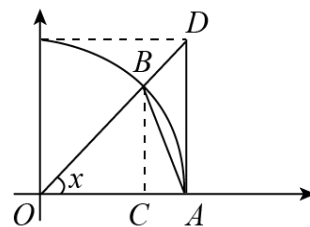
$$\because \cos x, \frac{\sin x}{x}, 1 \text{ 均是偶函数, } \therefore \text{在} \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \text{ 均成立.}$$

$$\cos x \leq \frac{\sin x}{x} \leq 1 \quad (\text{运用夹逼准则})$$

$$\therefore \lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$$

$$(4) \lim_{x \rightarrow \infty} \frac{\sin x}{x} = 0 \quad \text{依据定理 2.}$$

$$\text{作业: } \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}$$



1.6 函数连续

函数连续的定义:

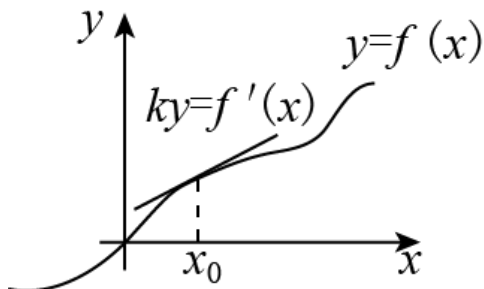
设函数 $y = f(x)$ 在点 x_0 的某一邻域内有定义, $\lim_{x \rightarrow x_0} f(x) = f(x_0)$, 则函数

$f(x)$ 在 x_0 处连续。

2. 导数

2.1 导数定义

导数: 变化快慢
微分: 变化程度



$$K = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

导数的定义:

$y = f(x)$ 在点 x_0 的某个邻域内有定义, 若 $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$ 存在, 则

称函数 $f(x)$ 在 x_0 处可导, 并称这个极限为函数 $y = f(x)$ 在点 x_0 处的导数。

例题: 求证 $y = \sin x$

$$\begin{aligned} \text{解: } f'(x) &= \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{2 \cos\left(x + \frac{1}{2} \Delta x\right) \cdot \sin \frac{\Delta x}{2}}{\Delta x} = \lim_{\Delta x \rightarrow 0} \cos\left(x + \frac{1}{2} \Delta x\right) \cdot \frac{\sin \frac{\Delta x}{2}}{\frac{\Delta x}{2}} \\ &= \cos x \end{aligned}$$

同理 $(\cos x)' = -\sin x$

$$\textcircled{1} \sin\left(x + h\right) = \sin\left(x + \frac{h}{2} + \frac{h}{2}\right) \quad \sin x = \sin\left(x + \frac{h}{2} - \frac{h}{2}\right)$$

$$\sin(a + \beta) = \sin a \cos \beta + \cos a \sin \beta$$

作业: $y = \cos x$ 的导数

导数的表达形式: $\frac{dy}{dx}$. $f'(x)$. y' . $\frac{df(x)}{dx}$

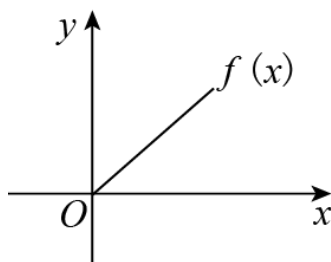
导数的几何意义: 切线的斜率

★可导必然连续, 连续未必可导.

例: $\text{ReLU}(x) = \max(x, 0)$

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

在 $x = 0$ 处连续但不可导。



2.2 导数的四则运算

$$(1) (au(x) \pm \beta v(x))' = au'(x) \pm \beta v'(x)$$

$$(2) [u(x) \cdot v(x)]' = u'(x)v(x) + u(x) \cdot v'(x)$$

$$(3) \left[\frac{u(x)}{v(x)} \right]' = \frac{u'(x) \cdot v(x) - u(x) \cdot v'(x)}{v^2(x)} (v(x) \neq 0)$$

$$\begin{aligned} (2) \text{ 证明: } f'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{u(x+h)v(x+h) - u(x)v(x)}{h} \\ &= \lim_{h \rightarrow 0} \left[\frac{u(x+h) - u(x)}{h} \cdot v(x+h) + u(x) \cdot \frac{v(x+h) - v(x)}{h} \right] \\ &= u'(x)v(x) + u(x)v'(x) \end{aligned}$$

2.3 高阶导数

表达形式: $f''(x_0)$, $y''|_{x=x_0}$; $\frac{d^2 y}{dx^2} \Big|_{x=x_0}$; $\frac{d^2 f(x)}{dx^2} \Big|_{x=x_0}$

例: $S = S(t)$

$$\text{速度: } v = \frac{ds}{dt} \quad \text{加速度: } a = \frac{dv}{dt}$$

$$a = \frac{dv}{dt} = \frac{d}{dt} \left(\frac{ds}{dt} \right) = \frac{d^2 s}{dt^2}$$

应用: 可判断凸凹性.

举例: 控制了房价过快上涨的趋势.

问题: $f(x) = \begin{cases} 6x^2, & x \geq 0 \\ 12x^2, & x < 0 \end{cases}$ 几阶可导?

2.4 链式求导 (复合函数求导)

(复合函数的求导链式规则, 简称链规则) 设 $y = y(u)$ 在点 u_0 处可导, $u = u(x)$

在点 x_0 处可导, $u_0 = u(x_0)$, 则复合函数 $y = y(u(x))$ 在点 x_0 处可导, 且

$$(y(u(x)))' \Big|_{x=x_0} = y'(u_0)u'(x_0), \text{ 或 } y'_x \Big|_{x=x_0} = y'_u \Big|_{u=u_0} \bullet u'_x \Big|_{x=x_0}.$$

证明:

$$\text{令 } F(\Delta u) = \begin{cases} \frac{y(u_0 + \Delta u) - y(u_0)}{\Delta u}, & \Delta u \neq 0, \\ y'(u_0), & \Delta u = 0, \end{cases}$$

则 $F(\Delta u)$ 在 $\Delta u = 0$ 处连续. 此外, 由 $u = u(x)$ 在点 x_0 处可导知,

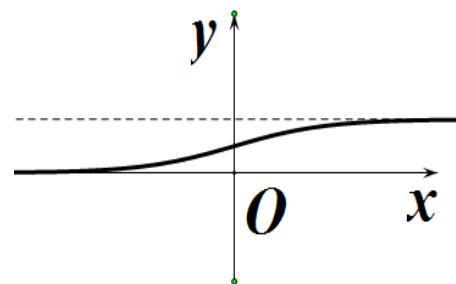
$\Delta u \rightarrow 0 (\Delta x \rightarrow 0)$. 于是:

$$\begin{aligned} (y(u(x)))' \Big|_{x=x_0} &= \lim_{\Delta x \rightarrow 0} \frac{y(u(x_0 + \Delta x)) - y(u(x_0))}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} F(\Delta u) \bullet \frac{\Delta u}{\Delta x} = y'(u_0)u'(x_0). \end{aligned}$$

2.5 Sigmoid 函数求导与梯度消失

例: $S(x) = \frac{1}{1 + e^{-x}}$

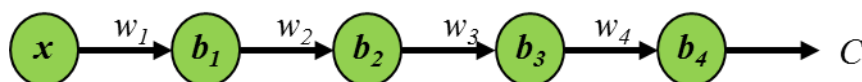
$$\begin{aligned} S'(x) &= \frac{-(-e^{-x})}{(1 + e^{-x})^2} = \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{1 + e^{-x}} \cdot \frac{1 + e^{-x} - 1}{1 + e^{-x}} \\ &= S(x) \cdot [1 - S(x)] < 1 \end{aligned}$$



为什么 Sigmoid 函数会导致梯度消失?

Sigmoid 激活函数“糟糕”的解析性质:

多层网络示例:



利用 BP 算法计算 b_1 的梯度:

$$\frac{\partial C}{\partial b_1} = \sigma'(b_1)w_2\sigma'(b_2)w_3\sigma'(b_3)w_4\sigma'(b_4)\frac{\partial C}{\partial \hat{b}_4}$$

Sigmoid 激活函数梯度: $\sigma'(x) = \sigma(x)(1 - \sigma(x)) \in \left(0, \frac{1}{4}\right]$

由此可得: $\frac{\partial C}{\partial b_1} \leq \left(\frac{1}{4}\right)^4 w_2 w_3 w_4 \frac{\partial C}{\partial \hat{b}_4}$

2.5 偏导数和梯度

偏导数定义:

设函数 $z = f(x, y)$ 在点 (x_0, y_0) 的某一邻域内有定义, 当 y 固定在 y_0 而 x 在 x_0 处有增量 Δx 时, 相应的函数有增量

$$f(x_0 + \Delta x, y_0) - f(x_0, y_0),$$

如果

$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x} \quad (2-1)$$

存在, 那么称此极限为函数 $z = f(x, y)$ 在点 (x_0, y_0) 处对 x 的偏导数, 记作

$$\left. \frac{\partial z}{\partial x} \right|_{\substack{x=x_0 \\ y=y_0}}, \left. \frac{\partial f}{\partial x} \right|_{\substack{x=x_0 \\ y=y_0}}, z_x \bigg|_{\substack{x=x_0 \\ y=y_0}} \text{ 或 } f_x(x_0, y_0). \quad (1)$$

例如, 极限 (2-1) 可以表为

$$f_x(x_0, y_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}. \quad (2-2)$$

类似地, 函数 $z = f(x, y)$ 在点 (x_0, y_0) 处对 y 的偏导数定义为

$$\lim_{\Delta y \rightarrow 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}, \quad (2-3)$$

记作

$$\left. \frac{\partial z}{\partial y} \right|_{\substack{x=x_0 \\ y=y_0}}, \left. \frac{\partial f}{\partial y} \right|_{\substack{x=x_0 \\ y=y_0}}, z_y \bigg|_{\substack{x=x_0 \\ y=y_0}} \text{ 或 } f_y(x_0, y_0).$$

如果函数 $z = f(x, y)$ 在区域 D 内每一点 (x, y) 处对 x 的偏导数都存在, 那么这个偏导数就是 x, y 的函数, 它就称为函数 $z = f(x, y)$ 对自变量 x 的偏导函数, 记作

$$\frac{\partial z}{\partial x}, \frac{\partial f}{\partial x}, z_x \text{ 或 } f_x(x, y).$$

类似地, 可以定义函数 $z = f(x, y)$ 对自变量 y 的偏导函数, 记作

$$\frac{\partial z}{\partial y}, \frac{\partial f}{\partial y}, z_y \text{ 或 } f_y(x, y).$$

由偏导数的概念可知, $f(x, y)$ 在点 (x_0, y_0) 处对 x 的偏导数 $f_x(x_0, y_0)$ 显然就是

偏导函数 $f_x(x, y)$ 在点 (x_0, y_0) 处的函数值； $f_x(x_0, y_0)$ 就是偏导函数 $f_x(x, y)$ 在点 (x_0, y_0) 处的函数值. 就像一元函数的导函数一样，以后在不至于混淆的地方也把偏导函数简称为偏导数。

将偏导函数写成向量形式就表示梯度函数。

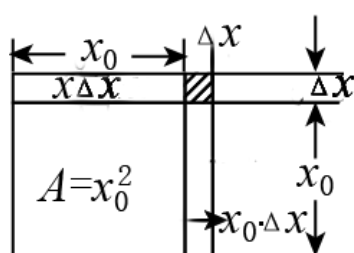
$$\nabla F(x, y, z) = \left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z} \right)$$

$$F(x, y, z) = x + y^2 + z^3$$

$$\nabla F(x, y, z) = (1, 2y, 3z^2)$$

3. 微分

3.1 微分的定义



$$S(x_0) = x_0^2, \quad S(x_0 + \Delta x) = (x_0 + \Delta x)^2$$

$$\Delta S = S(x_0 + \Delta x) - S(x_0)$$

$$= (x_0 + \Delta x)^2 - x_0^2$$

$$= 2x_0\Delta x + (\Delta x)^2$$

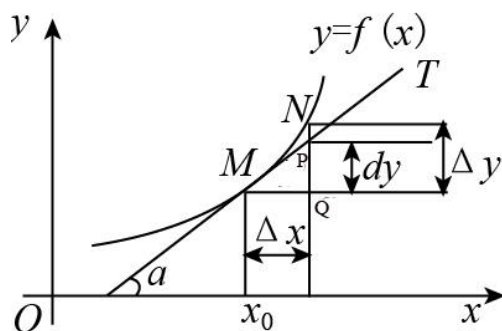
$$\Delta S \approx 2x_0 \cdot \Delta x \text{ (约去 } \Delta x \text{ 的高阶无穷小)}$$

微分的定义：

如果 $y = f(x)$ 在 x_0 的增量可表示为 $\Delta y = f(x_0 + \Delta x) - f(x_0) = A\Delta x + o(\Delta x)$ ， A 是与 x_0 有关的而不依赖于 Δx 的常数， $A \cdot \Delta x$ 称为函数 $y = f(x)$ 在点 x_0 相应于自变量的增量 Δx 的微分，记为 dy ，即 $dy = A \cdot \Delta x$ 。

可微的充要条件：① $f(x)$ 在 x_0 点处可导② $A = f'(x_0)$ 即 $dy = f'(x_0) \cdot \Delta x$

几何意义：切线纵坐标的增量



3.2 拉格朗日中值定理

若 $f(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 内可导, 则 $\exists \xi \in (a, b)$ 使得

$$f'(\xi) = \frac{f(b) - f(a)}{b - a} \quad (\text{利用罗尔定理证明})$$

3.3 泰勒公式

若 $f(x)$ 在 (a, b) 内有 $(n+1)$ 阶导数, 那么对 $x \in (a, b)$ 有

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \cdots + R_n(x)$$

其中, $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - x_0)^{n+1}$. ξ 是 x_0 与 x 之间的某值, $R_n(x)$ 称为拉格朗日

余项。

4. 积分

4.1 不定积分

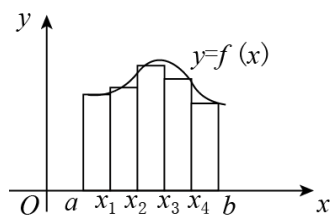
若: $F'(x) = f(x)$ 则 $\int f(x) dx = F(x) + C$.

$$\left. \begin{array}{l} F'(x) = (?) \\ (?)' = f(x) \end{array} \right\} \text{可逆运算}$$

$$\text{例题: } \int e^x dx = e^x + C, \quad \int k dx = kx + C, \quad \int x^u dx = \frac{1}{u+1} x^{u+1} + C.$$

4.2 定积分

求面积:



$$A = \lim_{\lambda \rightarrow 0} \sum_{i=1}^n f(\xi_i) \cdot \Delta x_i$$

$$\text{其中 } \lambda = \max \{ \Delta x_1, \Delta x_2, \dots, \Delta x_n \}$$

注意: $f(x) > 0$ $\int_a^b f(x) dx = A$ 表示曲线面积.

$f(x) < 0$ $\int_a^b f(x) dx = A$ 表示曲线面积的负值.

(1) 积分中值定理: 若 $f(x)$ 在 $[a, b]$ 上连续, 则 $[a, b]$ 上至少存在点 ξ . 使得

$$\int_a^b f(x) dx = f(\xi)(b-a) \quad (a \leq \xi \leq b).$$

(2) 定理 1: 若 $f(x)$ 在 $[a, b]$ 上连续, 那么积分上限函数 $\Phi(x) = \int_a^x f(t) dt$ 可导.

$$\text{导数为: } \Phi'(x) = \frac{d}{dx} \int_a^x f(t) dt = f(x) \quad (a \leq x \leq b)$$

(3) 定理 2: 由定理 1 可引出, 若 $f(x)$ 在 $[a, b]$ 上连续, 则积分上限函数 $\int_a^x f(t) dt$ 就是它的一个原函数.

(4) 牛顿—莱布尼茨公式: 若 $f(x)$ 在 $[a, b]$ 上连续, 函数 $F(x)$ 在 $[a, b]$ 上是 $f(x)$ 的一个原函数, 则: $\int_a^b f(x) dx = F(b) - F(a)$

(5) 牛顿—莱布尼茨公式的证明:

根据定理 2

$$\because F(x) \text{ 与 } \Phi(x) = \int_a^x f(t) dt \text{ 都是 } f(x) \text{ 原函数}$$

$$\therefore \Phi(x) = F(x) + C$$

$$\text{令 } x = a \quad \Phi(a) = F(a) + C$$

$$\therefore \int_a^x f(t) dt = F(x) + C = F(x) - F(a)$$

$$\text{令 } x = b \quad \text{则 } \int_a^b f(t) dt = F(b) - F(a)$$

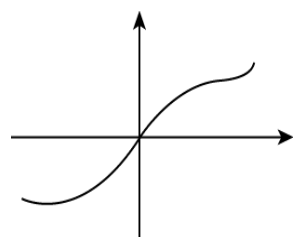
5. 单调与凸函数

5.1 单调增与单调减

定理: 设函数 $y = f(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 内可导.

(1) 如果在 (a,b) 内 $f'(x) \geq 0$, 且等号仅在有限多个点处成立, 那么函数 $y = f(x)$ 在 $[a,b]$ 上单调增加;

(2) 如果在 (a,b) 内 $f'(x) \leq 0$, 且等号仅在有限多个点处成立, 那么函数 $y = f(x)$ 在 $[a,b]$ 上单调减少.



判别: $f(x)$ 在 $[a,b]$ 上连续, 在 (a,b) 上可导.

① 对 $\forall x \in (a,b)$ $f'(x) > 0$ 则 $[a,b]$ 上单调增加

② 对 $\forall x \in (a,b)$ $f'(x) < 0$ 则 $[a,b]$ 上单调减少

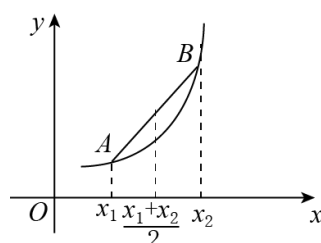
证明: 在 $[a,b]$ 上取 x_1, x_2 其中 $x_1 < x_2$, 在 $[x_1, x_2]$ 上应用拉格朗日中值定理:

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) \quad (x_1 < \xi < x_2),$$

若 $f'(\xi) > 0$, 则 $f(x_2) - f(x_1) > 0$; 若 $f'(\xi) < 0$, 则 $f(x_2) - f(x_1) < 0$.

5.2 凸函数判定

凸函数: 凹弧 (下凸的)

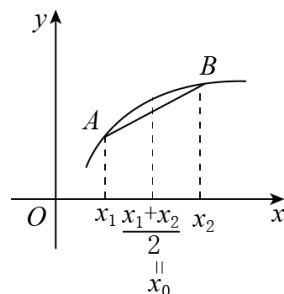


定义: $f(x)$ 在区间 I 连续, $x_1, x_2 \in I$.

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2}$$

判别: $f''(x) \geq 0$ 或 $f'(x)$ 在 I 内单调增加

凹函数: 凸弧 (上凸的)



定义: $f(x)$ 在区间 I 连续, $x_1, x_2 \in I$

$$f\left(\frac{x_1 + x_2}{2}\right) \geq \frac{f(x_1) + f(x_2)}{2}$$

判别: $f'(x)$ 在 I 内单调减少. $f''(x) \leq 0$.

证明: 根据微分中值定理:

取: $x_1 < \xi_1 < x_0 < \xi_2 < x_2$

$$\begin{cases} f(x_1) = f(x_0) + f'(\xi_1) \cdot (x_1 - x_0) & (x_1 < \xi_1 < x_0) \\ f(x_2) = f(x_0) + f'(\xi_2) \cdot (x_2 - x_0) & (x_0 < \xi_2 < x_2) \end{cases}$$
$$\begin{aligned} \frac{f(x_1) + f(x_2)}{2} &= f(x_0) + f'(\xi_1) \frac{(x_1 - x_0)}{2} + f'(\xi_2) \frac{(x_2 - x_0)}{2} \\ &= f(x_0) + \frac{1}{4}(x_2 - x_1)[f'(\xi_2) - f'(\xi_1)] \end{aligned}$$

其中: $f(x_0) = f\left(\frac{x_1 + x_2}{2}\right); (x_2 - x_1) > 0;$

$[f'(\xi_2) - f'(\xi_1)]$ 的值取决于 $f'(x)$ 单调性.

大作业: Softmax Loss 的梯度推导

高等数学作业

姓名： 得分：

1. $\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}$

2. 求 $y = \cos x$ 的导数

3. $f(x) = \begin{cases} 6x^2 & x \geq 0 \\ 12x^2 & x < 0 \end{cases}$ 几阶可导?

4. 证明中值定理：若 $f(x)$ 在 $[a, b]$ 上连续，在 (a, b) 内可导. 则 $\exists \xi \in (a, b)$ 使得

$$f'(\xi) = \frac{f(b) - f(a)}{b - a} \quad (\text{提示：罗尔定理})$$

5. 推导 Softmax Loss