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$D \subset R$,则称映射$f : D \to R$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

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

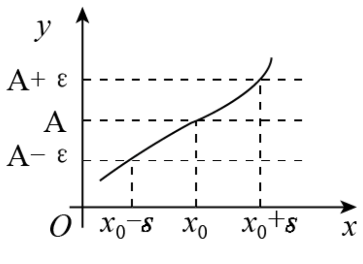
函数极限通常写为：$\lim\limits_{x \to x_0} f\left(x\right) = A$或$f\left(x\right) \to A$（当$x \to x_0$）

函数极限的定义：

如果$\forall \varepsilon > 0$，$\exists \delta > 0$，当$0 < \left|x - x_0\right| < \delta$时，总有$\left|f\left(x\right) - A\right| < \varepsilon$则$\lim\limits_{x \to x_0} f\left(x\right) = A$

等价于$\exists \delta > 0$，当$x \in \mathring{\bigcup}\left(x_0, \delta\right)$时总有$f\left(x\right) \in \bigcup\left(A, \varepsilon\right)$则$\lim\limits_{x \to x_0} f\left(x\right) = A$.

$x \in \mathring{\bigcup}\left(x_0, \delta\right)$的含义：点的$\delta$邻域$U\left(\alpha, \delta\right) = \left\{x \middle| \left|\left(x - \alpha\right)\right| < \delta\right\}$



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

$x \to x_0$ (或 $x \to \infty$) 时的无穷小.

特别地，以零为极限的数列 $\{x_n\}$ 称为 $n \to \infty$ 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 $x \to x_0$ (或 $x \to \infty$) 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) $\lim\left[f(x) \pm g(x)\right]=\lim f(x) \pm \lim g(x)=A \pm B$;

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

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

定理 4：设有数列 $\{x_n\}$ 和 $\{y_n\}$. 如果 $\lim\limits_{n \to \infty} x_n=A, \quad \lim\limits_{n \to \infty} y_n=B$, 那么

(1) $\lim\limits_{n \to \infty}(x_n \pm y_n)=A \pm B$;

(2) $\lim\limits_{n \to \infty}(x_n \cdot y_n)=A \cdot B$;

(3) 当 $y_n \neq 0 \ (n=1,2,\cdots)$ 且 $B \neq 0$ 时, $\lim\limits_{n \to \infty} \dfrac{x_n}{y_n}=\dfrac{A}{B}$.

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 $y=f\left[g(x)\right]$ 是由函数 $u=g(x)$ 与函数 $y=f(u)$ 复合而成， $f\left[g(x)\right]$ 在点 x_0 的某去心领域内有定义，若 $\lim\limits_{x \to x_0} g(x)=u_0, \lim\limits_{u \to u_0} f(u)=A$, 且存在 $\delta_0>0$, 当 $x \in \overset{o}{U}(x_0, \delta_0)$ 时，有 $g(x) \neq u_0$, 则

$\lim\limits_{x\to x_0} f\left[g(x)\right]=\lim\limits_{u\to u_0} f(u)=A.$

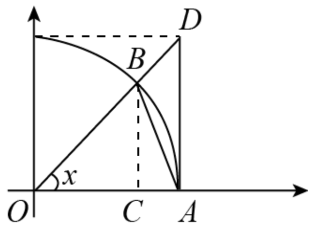
1.5 函数极限求解例题

（1）$\lim\limits_{x\to 1}\frac{x^2-1}{x-1}=\lim\limits_{x\to 1}\left(x+1\right)=2$

（2）$\lim\limits_{x\to\infty}\frac{1}{x}=0$

（3）$\lim\limits_{x\to 0}\frac{\sin x}{x}=1$

题（3）解：



$\because S_{\Delta}AOB\le S_{扇形}AOB\le S_{\Delta}AOD$

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

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

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

$\therefore \lim\limits_{x\to 0}\frac{\sin x}{x}=1$

（4）$\lim\limits_{x\to\infty}\frac{\sin x}{x}=0$ 依据定理 2.

作业： $\lim\limits_{x\to 0}\frac{1-\cos x}{x^2}$

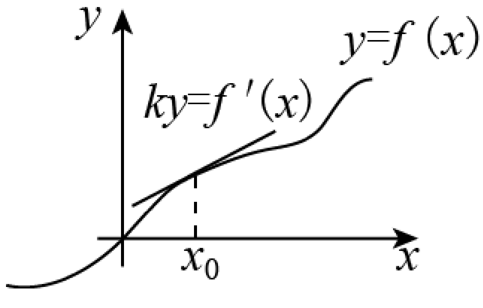
1.6 函数连续

函数连续的定义：

设函数 $y=f\left(x\right)$ 在点 x_0 的某一邻域内有定义，$\lim\limits_{x\to x_0} f\left(x\right)=f\left(x_0\right)$，则函数 $f\left(x\right)$ 在 x_0 处连续。

2. 导数

2.1 导数定义



$\begin{cases}\text{导数：变化快慢}\\ \text{微分：变化程度}\end{cases}$

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

导数的定义：

$y=f(x)$在点x_0的某个邻域内有定义，若$\lim\limits_{x\to x_0}\dfrac{f(x)-f(x_0)}{x-x_0}=\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}$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$y=\sin x$

解：$$f'(x)=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{\sin(x+\Delta x)-\sin x}{\Delta x}$$

$$=\lim_{\Delta x\to 0}\frac{2\cos\left(x+\frac{1}{2}\Delta x\right)\cdot\sin\frac{\Delta x}{2}}{\Delta x}=\lim_{\Delta x\to 0}\cos\left(x+\frac{1}{2}\Delta x\right)\cdot\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}$$

$$=\cos x$$

同理$(\cos x)'=-\sin x$

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

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

作业： $y=\cos x$的导数

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

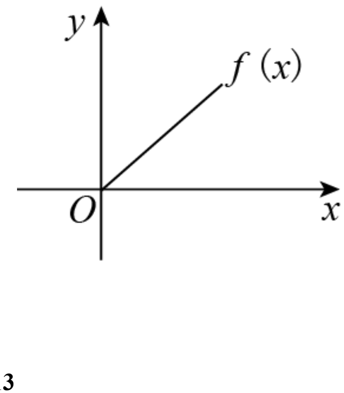
导数的几何意义：切线的斜率

☆可导必然连续，连续未必可导.

例：ReLU（x）= max(x, 0)

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

在$x=0$处连续但不可导。



2.2 导数的四则运算

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

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

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

（2）证明：$f'(x)=\lim\limits_{h\to 0}\dfrac{f(x+h)-f(x)}{h}$

$$=\lim_{h\to 0}\frac{u(x+h)v(x+h)-u(x)v(x)}{h}$$

$$=\lim_{h\to 0}\left[\frac{u(x+h)-u(x)}{h}\cdot v(x+h)+u(x)\cdot\frac{v(x+h)-v(x)}{h}\right]$$

$$=u'(x)v(x)+u(x)v'(x)$$

2.3 高阶导数

表达式： $f''(x_0)$，$y''\big|_{x=x_0}$；$\left.\dfrac{d^2y}{dx^2}\right|_{x=x_0}$；$\left.\dfrac{d^2f(x)}{dx^2}\right|_{x=x_0}$

例：$S=S(t)$

速度：$v=\dfrac{ds}{dt}$　　　加速度：$a=\dfrac{dv}{dt}$

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

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

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

证明：

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

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，$\Delta u\to 0(\Delta x\to 0)$. 于是：

$$(y(u(x)))'\big|_{x=x_0}=\lim_{\Delta x\to 0}\frac{y(u(x_0+\Delta x))-y(u(x_0))}{\Delta x}$$

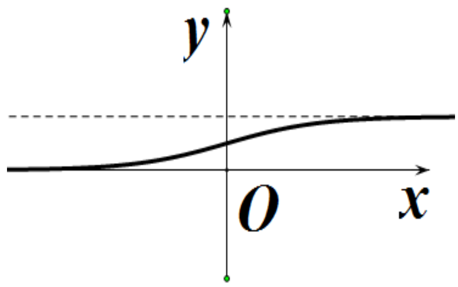
$$=\lim_{\Delta x\to 0}F(\Delta u)\cdot\frac{\Delta u}{\Delta x}=y'(u_0)u'(x_0).$$

2.5 Sigmoid 函数求导与梯度消失

例：$S(x)=\dfrac{1}{1+e^{-x}}$

$$S'(x)=\frac{-\left(-e^{-x}\right)}{\left(1+e^{-x}\right)^2}=\frac{e^{-x}}{\left(1+e^{-x}\right)^2}=\frac{1}{1+e^{-x}}\cdot\frac{1+e^{-x}-1}{1+e^{-x}}$$

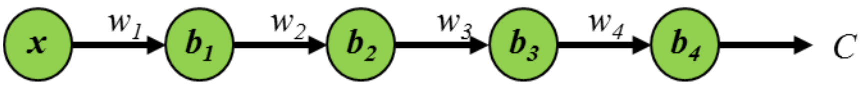
$$=S(x)\cdot\left[1-S(x)\right]\quad <1$$



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：$\sigma'(x)=\sigma(x)(1-\sigma(x))\in\left(0,\dfrac{1}{4}\right]$

由此可得：$\dfrac{\partial C}{\partial b_1}\le\left(\dfrac{1}{4}\right)^4 w_2w_3w_4\dfrac{\partial C}{\partial \hat{b}_4}$

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

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

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

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

例如，极限（2-1）可以表为

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

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

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

记作

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

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； $f_y(x_0,y_0)$ 就是偏导函数 $f_y(x,y)$ 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。

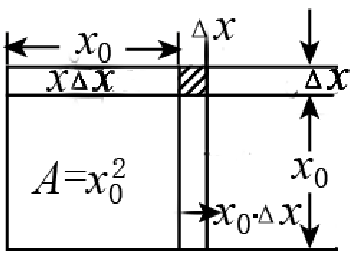
$$\nabla F(x,y,z)=(\frac{\partial F}{\partial x},\frac{\partial F}{\partial y},\frac{\partial F}{\partial z})$$

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

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

3. 微分

3.1 微分的定义



$$S\left(x_0\right)=x_0{}^2，\quad S\left(x_0+\Delta x\right)=\left(x_0+\Delta x\right)^2$$

$$\begin{aligned}\Delta S&=S\left(x_0+\Delta x\right)-S\left(x_0\right)\\&=\left(x_0+\Delta x\right)^2-x_0{}^2\\&=2x_0\Delta x+\left(\Delta x\right)^2\end{aligned}$$

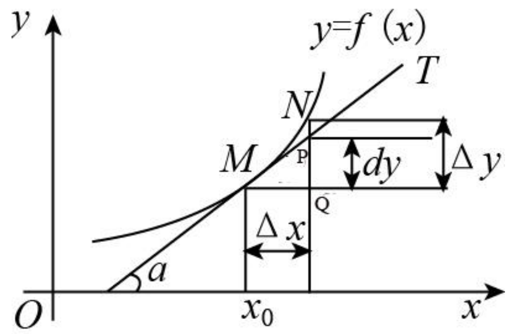
$\Delta S\approx 2x_0\cdot\Delta x$（约去$\Delta x$的高阶无穷小）

微分的定义：

如果 $y=f\left(x\right)$ 在 x_0 的增量可表示为 $\Delta y=f\left(x_0+\Delta x\right)-f\left(x_0\right)=A\Delta x+o\left(\Delta x\right)$，$A$ 是与 x_0 有关的而不依赖于 Δx 的常数，$A\cdot\Delta x$ 称为函数 $y=f\left(x\right)$ 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 $f\left(x\right)$ 在 x_0 点处可导② $A=f'\left(x_0\right)$ 即 $dy=f'\left(x_0\right)\cdot\Delta x$

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

$$f'(\xi)=\frac{f(b)-f(a)}{b-a}$$（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，$R_n(x)=\dfrac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{(n+1)}$. 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

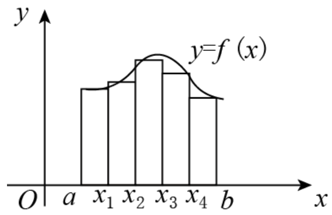
若：$F'(x)=f(x)$　则 $\int f(x)dx=F(x)+C$.

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

例题： $\int e^x dx=e^x+C$.　　$\int kdx=kx+C$.　　$\int x^u dx=\dfrac{1}{u+1}x^{u+1}+C$.

4.2 定积分

求面积:



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

其中 $\lambda=\max\left\{\Delta x_1,\Delta x_2\cdots\Delta x_n\right\}$

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

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

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

$$\int_a^b f\left(x\right)dx=f\left(\xi\right)\left(b-a\right)\ \left(a\le\xi\le b\right).$$

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

导数为：$\Phi'\left(x\right)=\dfrac{d}{dx}\int_a^x f\left(t\right)dt=f\left(x\right)\ \left(a\le x\le b\right)$

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

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

（5）牛顿—莱布尼茨公式的证明：

根据定理 2

$\because F\left(x\right)$ 与 $\Phi\left(x\right)=\int_a^x f\left(t\right)dt$ 都是 $f\left(x\right)$ 原函数

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

令 $x=a$　　$\Phi\left(a\right)=F\left(a\right)+C$

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

令 $x=b$　　则 $\int_a^b f\left(t\right)dt=F\left(b\right)-F\left(a\right)$

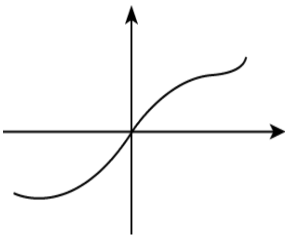
5. 单调与凸函数

5.1 单调增与单调减

定理： 设函数 $y=f(x)$ 在 $\left[a,b\right]$ 上连续，在 $\left(a,b\right)$ 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

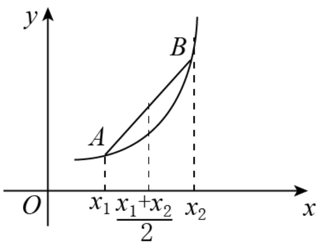
证明：在$[a,b]$上取x_1，x_2其中$x_1<x_2$，在$[x_1,x_2]$上应用**拉格朗日中值定理**:

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

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）

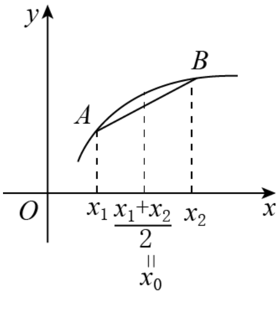


定义：$f(x)$在区间I连续，$x_1x_2\in I$.

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

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：$x_1 < \xi_1 < x_0 < \xi_2 < x_2$

$$\begin{cases} f\left(x_1\right) = f\left(x_0\right) + f'\left(\xi_1\right)\cdot\left(x_1 - x_0\right) & \left(x_1 < \xi_1 < x_0\right) \\ f\left(x_2\right) = f\left(x_0\right) + f'\left(\xi_2\right)\left(x_2 - x_0\right) & \left(x_0 < \xi_2 < x_2\right) \end{cases}$$

$$\frac{f\left(x_1\right) + f\left(x_2\right)}{2} = f\left(x_0\right) + f'\left(\xi_1\right)\frac{\left(x_1 - x_0\right)}{2} + f'\left(\xi_2\right)\frac{x_2 - x_0}{2}$$

$$= f\left(x_0\right) + \frac{1}{4}\left(x_2 - x_1\right)\left[f'\left(\xi_2\right) - f'\left(\xi_1\right)\right]$$

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

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

大作业：Softmax Loss 的梯度推导

高等数学作业

姓名：　　　得分：

1. $\lim\limits_{x \to 0} \dfrac{1-\cos x}{x^2}$

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

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

5. 推导 Softmax Loss