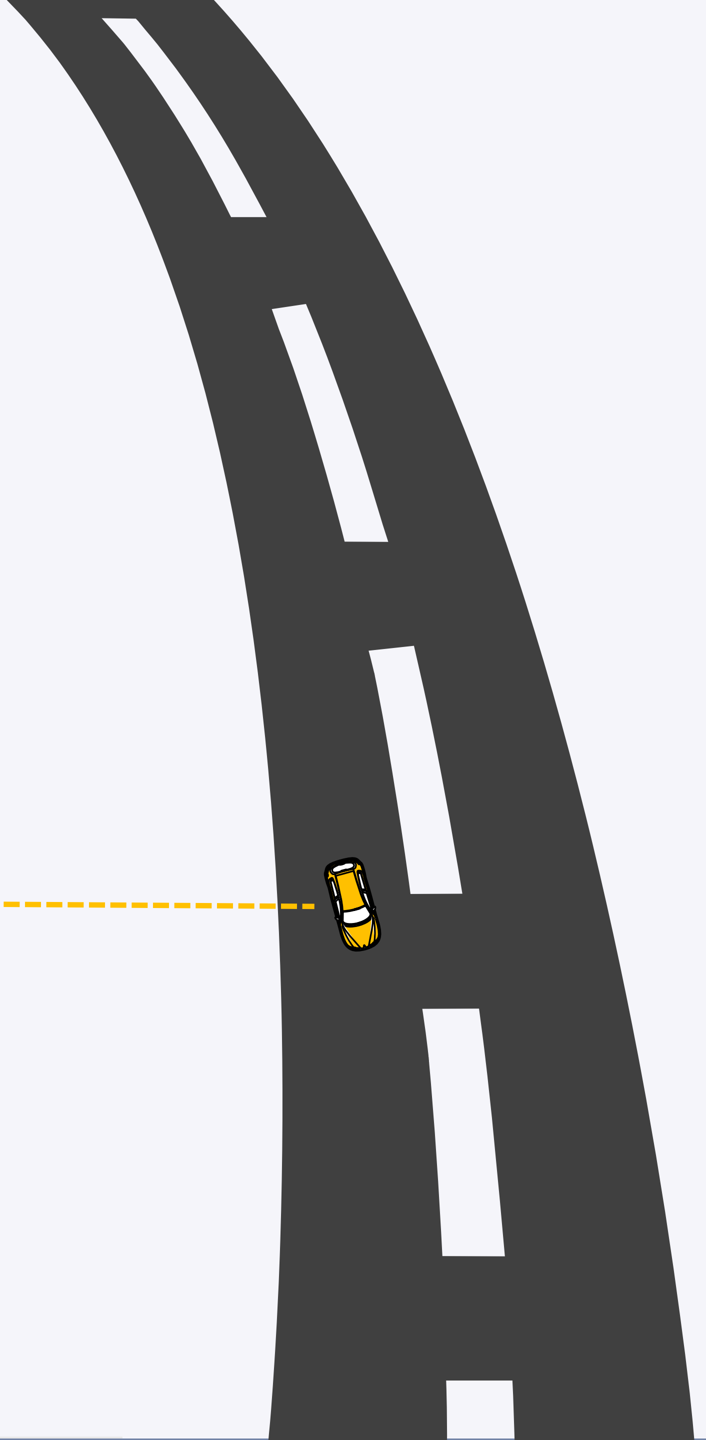
深度之眼
deepshare.net

数学基础—微积分

导师：Johnson

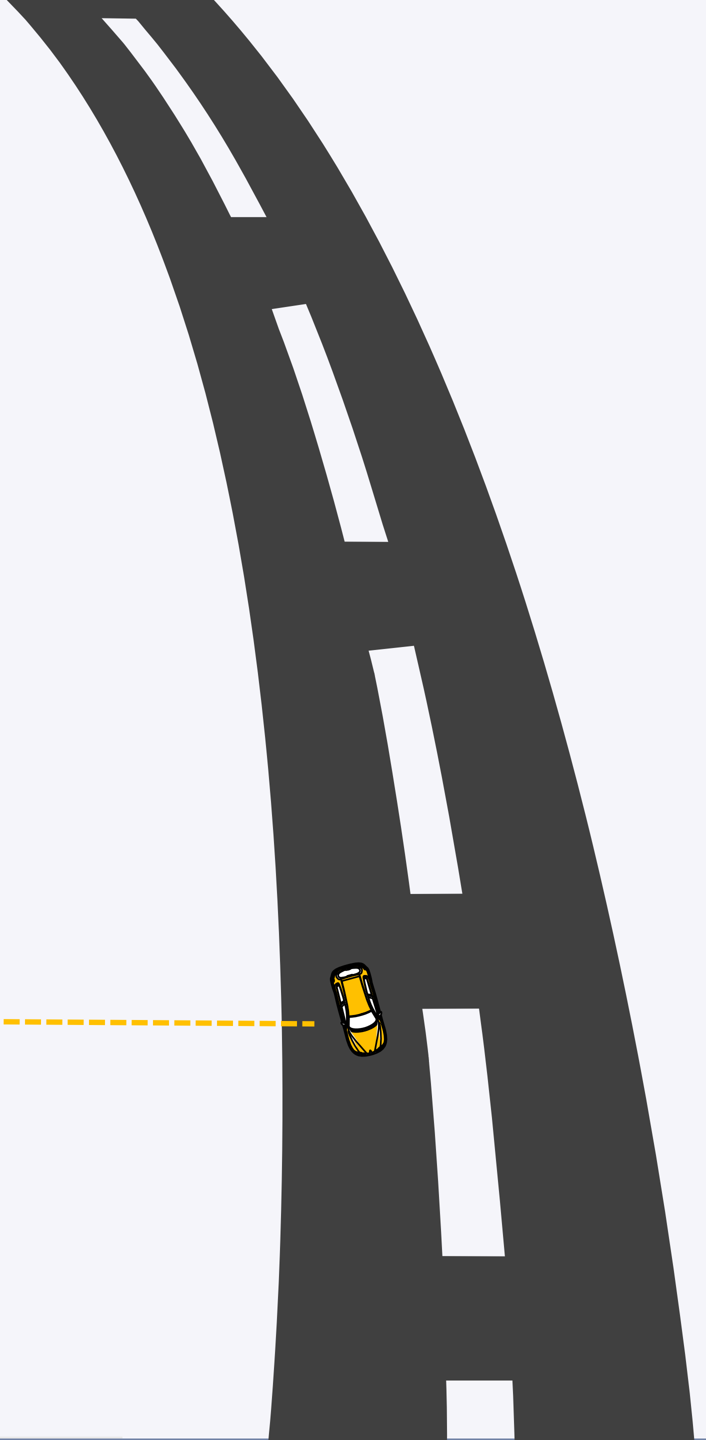
主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用



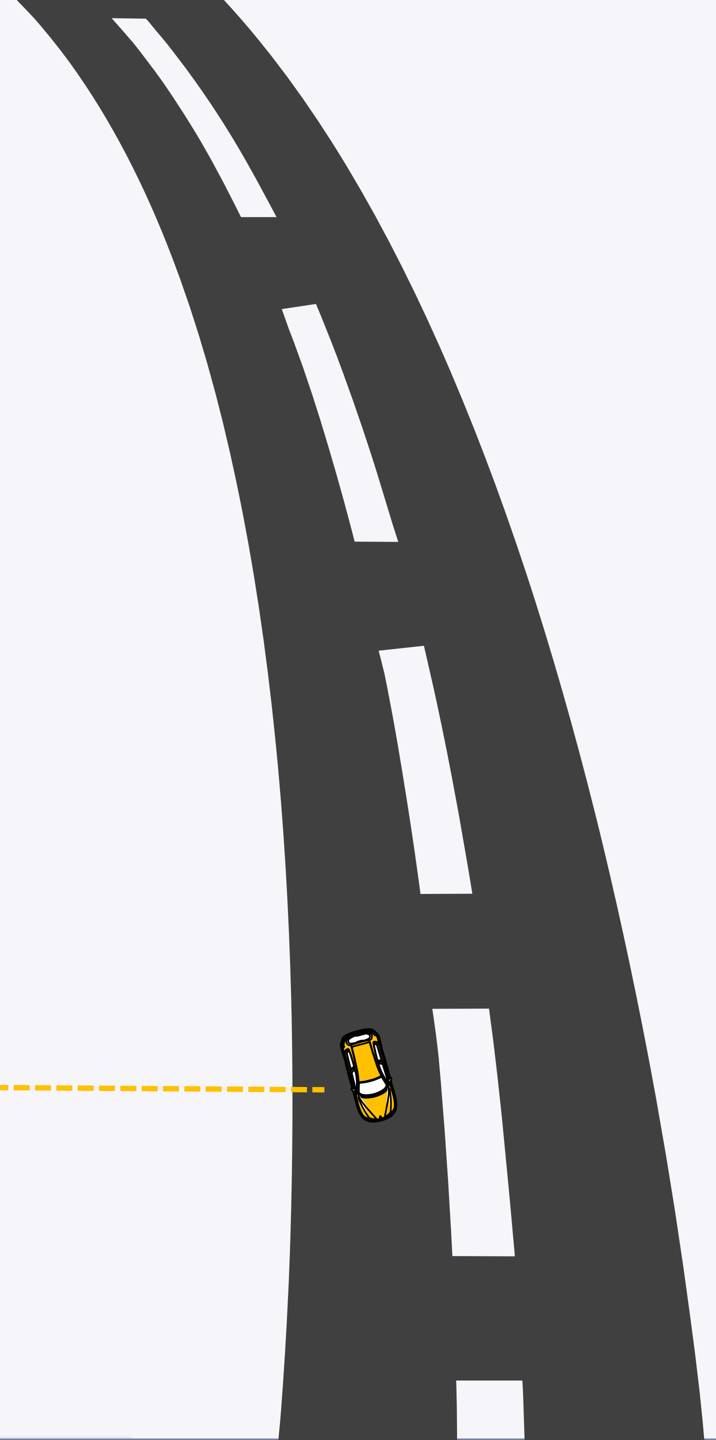
主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用



主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用

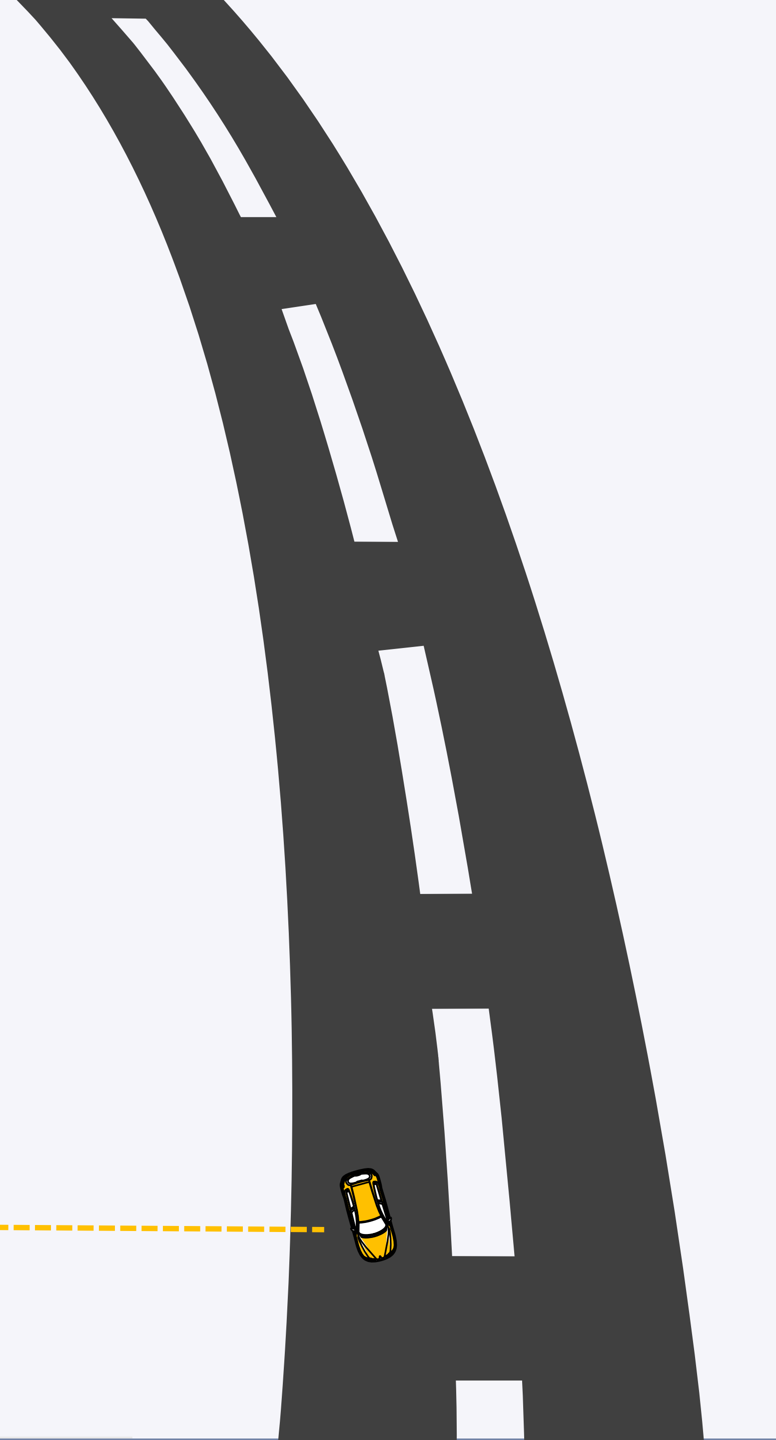


主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用

主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用



主要内容

- 微积分
  - P1:单变量微积分上(week3)
    - 1.导数的定义与意义
    - 2.常用函数的导数以及导数的常用公式，复合函数求导
    - 3.中值定理
    - 4.洛必达法则
    - 5.泰勒公式以及应用
  - P2:单变量微积分下(week3)
    - 1.函数的凹凸性
    - 2.函数的极值
    - 3.不定积分
    - 4.定积分
  - P3:多变量微积分(week3)
    - 1.偏导数
    - 2.多元复合函数的求导法则，链式求导法则
    - 3.方向导数与梯度及其应用
    - 4.多元函数泰勒公式与海森矩阵
    - 5.多元函数的极值
    - 6.矩阵的求导
    - 7.矩阵的求导在深度学习中的应用

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

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

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

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

偏导数

偏导数的概念还可推广到二元以上的函数
例如三元函数$u = f(x, y, z)$在点(x, y, z)处对x的偏导数定义为

$$f_x(x, y, z) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y, z) - f(x, y, z)}{\Delta x}$$

例1 求$z = x^2 + 3xy + y^2$在点$(1,2)$处的偏导数.

解 把y看做常量，得 $\frac{\partial z}{\partial x} = 2x + 3y$;

把x看做常量，得 $\frac{\partial z}{\partial y} = 3x + 2y$.

将(1,2)代入上面得结果，就得

$$\left.\frac{\partial z}{\partial x}\right|_{\substack{x=1\\y=2}} = 2 \cdot 1 + 3 \cdot 2 = 8, \left.\frac{\partial z}{\partial y}\right|_{\substack{x=1\\y=2}} = 3 \cdot 1 + 2 \cdot 2 = 7.$$

例2 求$z = x^2 \sin 2y$的偏导数.

解 $\frac{\partial z}{\partial x} = 2x \sin 2y$, $\frac{\partial z}{\partial y} = 2x^2 \cos 2y$.

偏导数

深度之眼
deepshare.net

设函数$z = f(x, y)$在区域D内具有偏导数

$$\frac{\partial z}{\partial x} = f_x(x, y), \frac{\partial z}{\partial y} = f_y(x, y),$$

于是在D内$f_x(x, y), f_y(x, y)$都是x, y的函数.如果这两个函数的偏导数也存在,那么称它们是函数$z = f(x, y)$的二阶偏导数.按照对变量求导次序的不同有下列四个二阶偏导数:

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x^2} = f_{xx}(x, y), \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial^2 z}{\partial x \partial y} = f_{xy}(x, y),$$

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y \partial x} = f_{yx}(x, y), \quad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial^2 z}{\partial y^2} = f_{yy}(x, y).$$

偏导数

例3 设$z = x^3y^2 - 3xy^3 - xy + 1$，求$\frac{\partial^2 z}{\partial x^2}$、$\frac{\partial^2 z}{\partial y \partial x}$、$\frac{\partial^2 z}{\partial x \partial y}$、$\frac{\partial^2 z}{\partial y^2}$及$\frac{\partial^3 z}{\partial x^3}$.

解 $\frac{\partial z}{\partial x} = 3x^2y^2 - 3y^3 - y$，$\frac{\partial z}{\partial y} = 2x^3y - 9xy^2 - x$；

$$\frac{\partial^2 z}{\partial x^2} = 6xy^2, \quad \frac{\partial^2 z}{\partial y \partial x} = 6x^2y - 9y^2 - 1;$$

$$\frac{\partial^2 z}{\partial x \partial y} = 6x^2y - 9y^2 - 1, \quad \frac{\partial^2 z}{\partial y^2} = 2x^3 - 18xy;$$

$$\frac{\partial^3 z}{\partial x^3} = 6y^2.$$

偏导数

定理 如果函数$z=f(x,y)$的两个二阶混合偏导数$\frac{\partial^2 z}{\partial y\partial x}$及$\frac{\partial^2 z}{\partial x\partial y}$在区间$D$内连续，那么在该区域内这两个二阶混合偏导数必相等.

$$z=x^3y^2-3xy^3-xy+1$$

$$\frac{\partial^2 z}{\partial y\partial x}=6x^2y-9y^2-1$$

$$\frac{\partial^2 z}{\partial x\partial y}=6x^2y-9y^2-1$$

多元复合函数求导法则，链式求导法则

1.一元函数与多元函数复合的情形

定理1 如果函数$u=\varphi(t)$及$v=\psi(t)$都在点t可导，函数$z=f(u,v)$在对应点(u,v)具有连续偏导数，那么复合函数$z=f[\varphi(t),\psi(t)]$在点t可导，且有

$$\frac{dz}{dt}=\frac{\partial z}{\partial u}\frac{du}{dt}+\frac{\partial z}{\partial v}\frac{dv}{dt}.$$

例1 z$=f(u,v)=uv$ $\quad u=e^t \quad v=\cos t$ 求$\frac{dz}{dt}$.

解

$$\frac{dz}{dt}=Z_u\frac{du}{dt}+Z_v\frac{dv}{dt}=ve^t+u(-\sin t)$$
$$=\cos t\cdot e^t-e^t\sin t=e^t(\cos t-\sin t)$$

链式求导很重要，

是深度学习的基础!!!

多元复合函数求导法则，链式求导法则

2.多元函数与多元函数复合的情形

定理2 如果函数$u=\varphi(x,y)$及$v=\psi(x,y)$都在点(x,y)具有对x及对y的偏导数，函数$z=f(u,v)$在对应点(u,v)具有连续偏导数，那么复合函数$z=f[\varphi(x,y),\psi(x,y)]$在点$(x,y)$的两个偏导数都存在，且有

$$\frac{\partial z}{\partial x}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial x}, \qquad (4-3)$$

$$\frac{\partial z}{\partial y}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial y}. \qquad (4-4)$$

例2 设$z=e^{u}\sin v$，而$u=xy$，$v=x+y$.求$\frac{\partial z}{\partial x}$和$\frac{\partial z}{\partial y}$.

解

$$\frac{\partial z}{\partial x}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial x}=e^{u}\sin v\cdot y+e^{u}\cos v\cdot 1$$
$$=e^{xy}[y\sin(x+y)+\cos(x+y)],$$

$$\frac{\partial z}{\partial y}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial y}=e^{u}\sin v\cdot x+e^{u}\cos v\cdot 1$$
$$=e^{xy}[x\sin(x+y)+\cos(x+y)].$$

多元复合函数求导法则，链式求导法则

例3 设$u = f(x,y,z) = e^{x^2+y^2+z^2}$，而$z = x^2 \sin y$．求$\frac{\partial u}{\partial x}$和$\frac{\partial u}{\partial y}$．

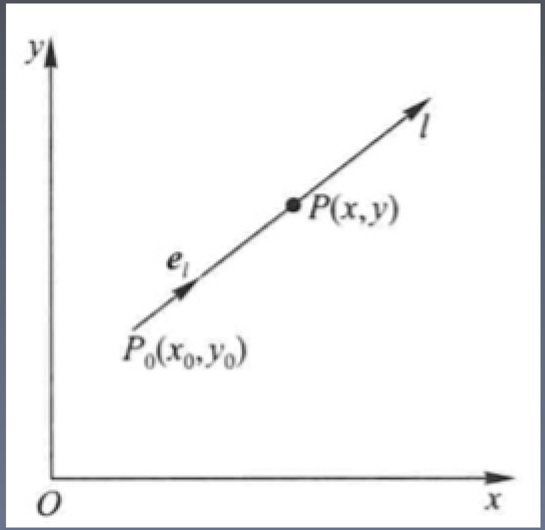
解
$$\frac{\partial u}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 2xe^{x^2+y^2+z^2} + 2ze^{x^2+y^2+z^2} \cdot 2x\sin y$$
$$= 2x(1 + 2x^2 \sin^2 y)e^{x^2+y^2+z^2}.$$

$$\frac{\partial u}{\partial y} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial y} = 2ye^{x^2+y^2+z^2} + 2ze^{x^2+y^2+z^2} \cdot x^2\cos y$$
$$= 2(y + x^4 \sin y \cos y)e^{x^2+y^2+x^4\sin^2 y}.$$

例4 设$z = f(u,v,t) = uv + \sin t$，而$u = e^t$，$v = \cos t$.求全导数$\frac{dz}{dt}$.

解
$$\frac{dz}{dt} = \frac{\partial f}{\partial u}\frac{du}{dt} + \frac{\partial f}{\partial v}\frac{dv}{dt} + \frac{\partial f}{\partial t} = ve^t - u\sin t + \cos t$$
$$= e^t\cos t - e^t\sin t + \cos t = e^t(\cos t - \sin t) + \cos t.$$

方向导数与梯度及其应用



$$\begin{cases} x = x_0 + t\cos\alpha, \\ y = y_0 + t\cos\beta \end{cases} (t \geq 0).$$

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = \lim_{t\to 0^+} \frac{f(x_0 + t\cos\alpha, y_0 + t\cos\beta) - f(x_0, y_0)}{t}.$$

方向导数

$$\lim_{t\to 0^+} \frac{f(x_0 + t\cos\alpha, y_0 + t\cos\beta) - f(x_0 + t\cos\alpha, y_0) + f(x_0 + t\cos\alpha, y_0) - f(x_0, y_0)}{t}$$

$$= \lim_{t\to 0^+} \frac{f(x_0 + t\cos\alpha, y_0 + t\cos\beta) - f(x_0 + t\cos\alpha, y_0)}{t\cos\beta}\cos\beta$$

$$+ \lim_{t\to 0^+} \frac{f(x_0 + t\cos\alpha, y_0) - f(x_0, y_0)}{t\cos\alpha}\cos\alpha.$$

方向导数与梯度及其应用

定理 如果函数$f(x,y)$在点$P_0(x_0,y_0)$可微分，那么函数在该点沿任一方向l的方向导数存在，且有

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)}=f_x(x_0,y_0)\cos\alpha+f_y(x_0,y_0)\cos\beta\,.$$

方向导数与梯度及其应用

与方向导数有关联的一个概念是函数的梯度.在二元函数的情形，设函数$f(x,y)$在平面区域D内具有一阶连续偏导数，则对于每一点$P_0(x_0,y_0)\in D$，都可定出一个向量

$$f_x(x_0,y_0)i+f_y(x_0,y_0)j,$$

这向量称为函数$f(x,y)$在点$P_0(x_0,y_0)$的梯度，记作$gradf(x_0,y_0)$或$\nabla f(x_0,y_0)$，即

$$gradf(x_0,y_0)=\nabla f(x_0,y_0)=f_x(x_0,y_0)i+f_y(x_0,y_0)j.$$

如果函数$f(x,y)$在点$P_0(x_0,y_0)$可微分，$e_l=(\cos\alpha,\cos\beta)$是与方向$l$同向的单位向量，

那么

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)}=f_x(x_0,y_0)\cos\alpha+f_y(x_0,y_0)\cos\beta$$

$$=gradf(x_0,y_0)\cdot e_l=|gradf(x_0,y_0)|\cos\theta$$

方向导数与梯度及其应用

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = f_x(x_0,y_0)\cos\alpha + f_y(x_0,y_0)\cos\beta$$
$$= gradf(x_0,y_0)\cdot e_l = |gradf(x_0,y_0)|\cos\theta$$

(1)当$\theta = 0$，即方向e_l与梯度$gradf(x_0,y_0)$的方向相同时，函数$f(x,y)$增加最快.此时，函数在这个方向的方向导数达到最大值，这个最大值就是梯度$gradf(x_0,y_0)$的模，即

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = |gradf(x_0,y_0)|.$$

这个结果也表示：函数$f(x,y)$在一点的梯度$gradf$是这样一个向量，它的方向是函数在这点的方向导数取得最大值的方向，它的模就等于方向导数的最大值.

梯度很重要，
是最优化的基础!!!

深度之眼
deepshare.net

方向导数与梯度及其应用

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = f_x(x_0,y_0)\cos\alpha + f_y(x_0,y_0)\cos\beta$$
$$= \mathrm{grad}f(x_0,y_0)\cdot e_l = |\mathrm{grad}f(x_0,y_0)|\cos\theta$$

(2)当$\theta = \pi$，即方向e_l与梯度$\mathrm{grad}f(x_0,y_0)$的方向相反时，函数$f(x,y)$减少最快.函数在这个方向的方向导数达到最小值，即

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = -|\mathrm{grad}f(x_0,y_0)|.$$

(3)当$\theta = \frac{\pi}{2}$，即方向e_l与梯度$\mathrm{grad}f(x_0,y_0)$的方向正交时，函数的变化率为零，即

$$\left.\frac{\partial f}{\partial l}\right|_{(x_0,y_0)} = |\mathrm{grad}f(x_0,y_0)|\cos\theta = 0.$$

梯度很重要，是最优化的基础!!!

方向导数与梯度及其应用

例1 设$f(x,y)=\frac{1}{2}(x^2+y^2), P_0(1,1)$，求
(1)$f(x,y)$在P_0处增加最快的方向以及$f(x,y)$沿这个方向的方向导数；
(2)$f(x,y)$在P_0处减少最快的方向以及$f(x,y)$沿这个方向的方向导数；
(3)$f(x,y)$在P_0处的变化率为零的方向.

解 (1)$f(x,y)$在P_0处沿$\nabla f(1,1)$的方向增加最快,

$$\nabla f(1,1)=(xi+yj)\Big|_{(1,1)}=i+j,$$

故所求方向可取为

$$n=\frac{\nabla f(1,1)}{|\nabla f(1,1)|}=\frac{1}{\sqrt{2}}i+\frac{1}{\sqrt{2}}j,$$

方向导数为

$$\frac{\partial f}{\partial n}\bigg|_{(1,1)}=|\nabla f(1,1)|=\sqrt{2}.$$

方向导数与梯度及其应用

深度之眼 deepshare.net

例1 设$f(x,y)=\frac{1}{2}(x^2+y^2)$, $P_0(1,1)$，求

(1)$f(x,y)$在P_0处增加最快的方向以及$f(x,y)$沿这个方向的方向导数;
(2)$f(x,y)$在P_0处减少最快的方向以及$f(x,y)$沿这个方向的方向导数;
(3)$f(x,y)$在P_0处的变化率为零的方向.

(2)$f(x,y)$在P_0处沿$-\nabla f(1,1)$的方向减少最快,

这方向可取为

$$n_1=-n=-\frac{1}{\sqrt{2}}i-\frac{1}{\sqrt{2}}j,$$

方向导数为

$$\left.\frac{\partial f}{\partial n_1}\right|_{(1,1)}=-|\nabla f(1,1)|=-\sqrt{2}.$$

方向导数与梯度及其应用

例1 设$f(x,y)=\frac{1}{2}(x^2+y^2)$, $P_0(1,1)$，求
(1)$f(x,y)$在P_0处增加最快的方向以及$f(x,y)$沿这个方向的方向导数;
(2)$f(x,y)$在P_0处减少最快的方向以及$f(x,y)$沿这个方向的方向导数;
(3)$f(x,y)$在P_0处的变化率为零的方向.

(3)$f(x,y)$在P_0处沿垂直于$\nabla f(1,1)$的方向变化率为零,
这方向是

$$n_2=-\frac{1}{\sqrt{2}}i+\frac{1}{\sqrt{2}}j \quad 或 \quad n_3=\frac{1}{\sqrt{2}}i-\frac{1}{\sqrt{2}}j.$$

方向导数与梯度及其应用

例2 设$f(x,y,z)=x^3-xy^2-z$, $P_0(1,1,0)$.问$f(x,y,z)$在P_0处沿什么方向变化最快，在这个方向的变化率是多少?

解 $\nabla f=f_x i+f_y j+f_z k=(3x^2-y^2)i-2xyj-k$, $\nabla f(1,1,0)=2i-2j-k$.

$f(x,y,z)$在P_0处沿$\nabla f(1,1,0)$的方向增加最快，沿$-\nabla f(1,1,0)$的方向减少最快，在这两个方向的变化率分别是

$$|\nabla f(1,1,0)|=\sqrt{2^2+(-2)^2+1}=3,$$
$$-|\nabla f(1,1,0)|=-3.$$

求极值梯度下降法

$$\min f(x)$$
$$x_{n+1}=x_n-\lambda\nabla f(x_n)$$

多元函数泰勒公式与海森矩阵

深度之眼
deepshare.net

定理 设$z=f(x,y)$在点(x_0,y_0)的某一邻域内连续且有$(n+1)$阶连续偏导数，(x_0+h,y_0+k)在此邻域内任一点，则有

$$
\begin{aligned}
&f(x_0+h,y_0+k)\\
&=f(x_0,y_0)+\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)f(x_0,y_0)+\frac{1}{2!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^2 f(x_0,y_0)+\cdots+\\
&\quad\frac{1}{n!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^n f(x_0,y_0)+\frac{1}{(n+1)!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{n+1} f(x_0+\theta h,y_0+\theta k)\qquad(0<\theta<1).
\end{aligned}
$$

其中记号

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)f(x_0,y_0)\text{表示}hf_x(x_0,y_0)+kf_y(x_0,y_0),$$

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^2 f(x_0,y_0)\text{表示}h^2f_{xx}(x_0,y_0)+2hkf_{xy}(x_0,y_0)+k^2f_{yy}(x_0,y_0)$$

一般地，记号

$$\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^m f(x_0,y_0)\text{表示}\sum_{p=0}^{m}C_m^p h^p k^{m-p}\left.\frac{\partial^m f}{\partial x^p\partial y^{m-p}}\right|_{(x_0,y_0)}.$$

海森矩阵(二维或高维)

深度之眼 deepshare.net

保留到二阶

$$f(x+\Delta x, y+\Delta y) = f(x,y) + f_x\Delta x + f_y\Delta y + f_{xx}\Delta x^2 + 2f_{xy}\Delta x\Delta y + f_{yy}\Delta y^2$$

$$= f(x,y) + \begin{pmatrix} f_x \\ f_y \end{pmatrix}^T \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} + (\Delta x, \Delta y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

掌握海森矩阵!!!

更一般地

$$f(x_1+\Delta x_1, \cdots, x_n+\Delta x_n)$$

$$= f(x_1,\cdots,x_n) + \begin{pmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{pmatrix}^T \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} + (\Delta x_1,\cdots,\Delta x_n)\begin{pmatrix} f_{x_1x_1} & \cdots & f_{x_1x_n} \\ \vdots & \ddots & \vdots \\ f_{x_nx_1} & \cdots & f_{x_nx_n} \end{pmatrix}\begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}$$

$$= f(x_1,\cdots,x_n) + \nabla f^T \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} + (\Delta x_1,\cdots,\Delta x_n)H\begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix}$$

多元函数的极值

定义 设函数$z = f(x, y)$的定义域为D，$P_0(x_0, y_0)$为D的内点.若存在P_0的某个邻域$U(P_0) \subset D$，使得对于该邻域内异于P_0的任何点(x, y)，都有

$$f(x, y) < f(x_0, y_0),$$

则称函数$f(x, y)$在点(x_0, y_0)有极大值(x_0, y_0)，点(x_0, y_0)称为函数$f(x, y)$的极大值点；若对于该邻域内异于P_0的任何点(x_0, y_0)，都有

$$f(x, y) > f(x_0, y_0),$$

则称函数$f(x, y)$在点(x_0, y_0)有极小值(x_0, y_0)，点(x_0, y_0)称为函数$f(x, y)$的极小值点.极大值与极小值统称为极值.使得函数取得极值的点称为极值点.

多元函数的极值

深度之眼 deepshare.net

定理1（必要条件） 设函数$z = f(x, y)$在点(x_0, y_0)具有偏导数，且在点(x_0, y_0)处有极值，则有

$$f_x(x_0, y_0) = 0, f_y(x_0, y_0) = 0.$$

定理2（充分条件） 设函数$z = f(x, y)$在点(x_0, y_0)的某邻域内连续且有一阶及二阶连续偏导数，又$f_x(x_0, y_0) = 0$，$f_y(x_0, y_0) = 0$，令

$$f_{xx}(x_0, y_0) = A, f_{xy}(x_0, y_0) = B, f_{yy}(x_0, y_0) = C,$$

则$f(x, y)$在(x_0, y_0)处是否取得极值的条件如下：

（1）$AC - B^2 > 0$时具有极值，且当$A < 0$时有极大值，当$A > 0$时有极小值；

（2）$AC - B^2 < 0$时没有极值；

（3）$AC - B^2 = 0$时可能有极值，也可能没有极值，还需另作讨论.

多元函数的极值

深度之眼
deepshare.net

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0, y_0) + \nabla f^T\Big|_{\substack{x=x_0\\y=y_0}} + (\Delta x, \Delta y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy}\end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} + \cdots$$

$$= f(x_0, y_0) + (\Delta x, \Delta y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy}\end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} + \cdots$$

①若H为正定，则$f(x_0+\Delta x, y_0+\Delta y) > f(x_0, y_0)$ 当$\Delta x, \Delta y$比较小时，

此时 $H = \begin{pmatrix} A & B \\ B & C\end{pmatrix}$ $|H-\lambda I| = \begin{vmatrix} A-\lambda & B \\ B & C-\lambda\end{vmatrix} = (\lambda - A)(\lambda - C) - B^2$

$$= \lambda^2 - (A+C)\lambda + AC - B^2 = 0.$$

②若H为负定，则$f(x_0+\Delta x, y_0+\Delta y) < f(x_0, y_0)$

$$\begin{cases}\lambda_1 + \lambda_2 = A + C < 0 \\ \lambda_1\lambda_2 = AC - B^2 > 0\end{cases} \Rightarrow A < 0, C < 0, AC > B^2$$

矩阵的求导

1. $f(x) = Ax$，则

$$\frac{\partial f(x)}{\partial x^T} = \frac{\partial (Ax)}{\partial x^T} = A$$

设 $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$ $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

$$f = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{pmatrix}$$

$$\frac{\partial f}{\partial X^T} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = A$$

矩阵的求导

2.$f(x) = x^T A x$，则

$$\frac{\partial f(x)}{\partial x} = \frac{\partial (x^T A x)}{\partial x} = Ax + A^T x$$

设 $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$$f(x) = (x_1\ x_2)\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a x_1^2 + (b+c)x_1 x_2 + d x_2^2$$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2ax_1 + (b+c)x_2 \\ (b+c)x_1 + 2dx_2 \end{pmatrix} = \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a & c \\ b & d \end{pmatrix}\right]\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (A + A^T)x$$

特别地当A为对称矩阵时 $\frac{\partial f}{\partial x} = 2Ax$

矩阵的求导

3. $f(x) = a^T x$，则

$$\frac{\partial a^T x}{\partial x} = \frac{\partial x^T a}{\partial x} = a$$

设 $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad f = a_1 x_1 + a_2 x_2$

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a \qquad \frac{\partial f}{\partial a} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x$$

$J = \|xa - y\|^2$ $\qquad x$为$N \times n, a$为$n \times 1,$

$J = (xa - y)^T (xa - y)$ $\qquad y$为$N \times 1$

$\quad = (a^T x^T - y^T)(xa - y)$

$\quad = a^T x^T x a - a^T x^T y - y^T x a + y^T y$

$$\frac{\partial (a^T x^T x a)}{\partial a} = 2x^T x a$$

$y^T x$为$1 \times n, a$为$n \times 1,$
$x^T y$为$n \times 1$

$$\frac{\partial (y^T x a)}{\partial a} = \frac{\partial [(x^T y)^T a]}{\partial a} = x^T y$$

$a^T x^T y = a^T (x^T y)$ $\qquad a$为$n \times 1 \; x^T y$为$n \times 1$

$\quad = (x^T y)^T a = y^T x a$

$$\frac{\partial J}{\partial a} = 2x^T x a - 2x^T y = 0$$

$x^T x a = x^T y$

矩阵的求导

4.$f(x)=x^TAy$，则

$$\frac{\partial x^TAy}{\partial x}=Ay$$

$$\frac{\partial x^TAy}{\partial A}=xy^T$$

5.$\frac{\partial\left(tr(zz^T)\right)}{\partial z}=\frac{\partial\left(tr(z^Tz)\right)}{\partial z}=2Z$　　Z为$m\times n$　z^T为$n\times m$

$$tr(ZZ^T)=\sum_{i=1}^{m}d_{ii}=\sum_{i=1}^{m}\sum_{j=1}^{n}z_{ij}z_{ji}^T=\sum_{i=1}^{m}\sum_{j=1}^{n}z_{ij}^2$$

$$\frac{\partial\left(tr(ZZ^T)\right)}{\partial Z}=\begin{pmatrix}2z_{11} & \cdots & 2z_{1n}\\ \vdots & \ddots & \vdots\\ 2z_{m1} & \cdots & 2z_{mn}\end{pmatrix}=2Z$$

矩阵的求导

$$\frac{\partial tr(A)}{\partial A} = I_{n\times n}$$

$$\frac{\partial tr(AB)}{\partial A} = B^T$$

$$tr(A) = \sum_{i=1}^{n} a_{ii} \quad A\text{为}n \times n$$

A为$m \times n$，
B为$n \times m$

$$\frac{\partial (trA)}{\partial A} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} = I$$

$$tr(AB) = \sum_{i=1}^{m} d_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji}$$

$$\frac{\partial (tr(AB))}{\partial A} = \begin{pmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1m} & \cdots & b_{nm} \end{pmatrix} = B^T$$

$$\frac{\partial |Z|}{\partial Z} = |Z|(Z^{-1})^T$$

$$\frac{\partial |Z|}{\partial Z_{ij}} = A_{ij}$$

$$|Z| = Z_{i1}A_{i1} + \cdots + Z_{ij}A_{ij} + \cdots + Z_{in}A_{in}$$

$$\frac{\partial |Z|}{\partial Z} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} = {Z^*}^T$$

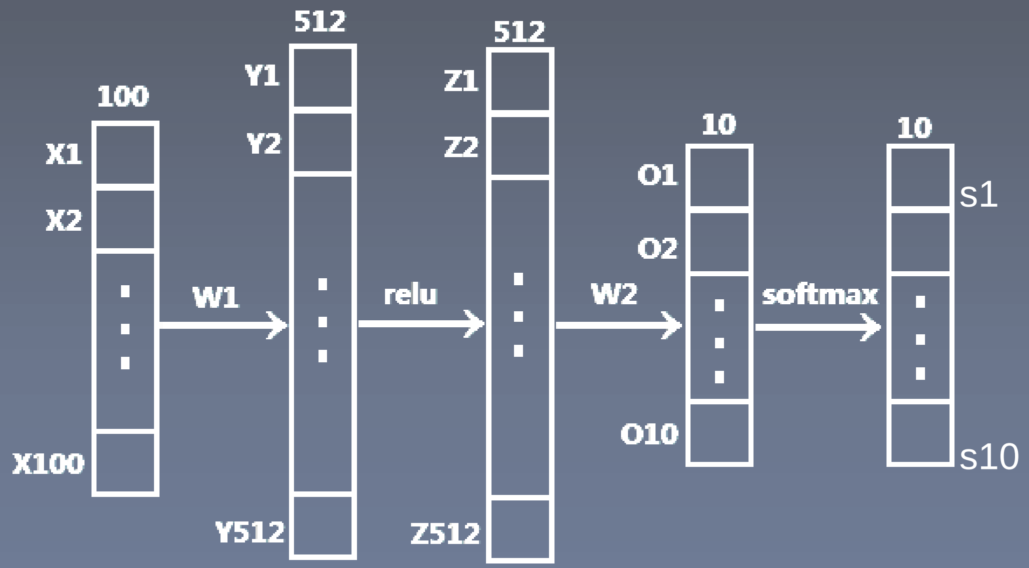
$$Z^*Z = ZZ^* = |Z|E$$

$$Z^* = Z^{-1}|Z|$$

$${Z^*}^T = |Z|(Z^{-1})^T$$

矩阵的求导在深度学习中的应用(中级)

深度之眼
deepshare.net



$W1$: $100 * 512$ 的矩阵

$W2$: $512 * 10$ 的矩阵

$$(x_1, x_2, \cdots, x_{100})w_1 = (y_1, y_2, \cdots, y_{512})$$

$$(z_1, z_2, \cdots, z_{512})w_2 = (o_1, o_2, \cdots, o_{10})$$

$$S_k = \frac{e^{o_k}}{\sum_{i=1}^{10} e^{o_i}}$$

$$J = -\ln S_k = J(w_1, w_2)$$

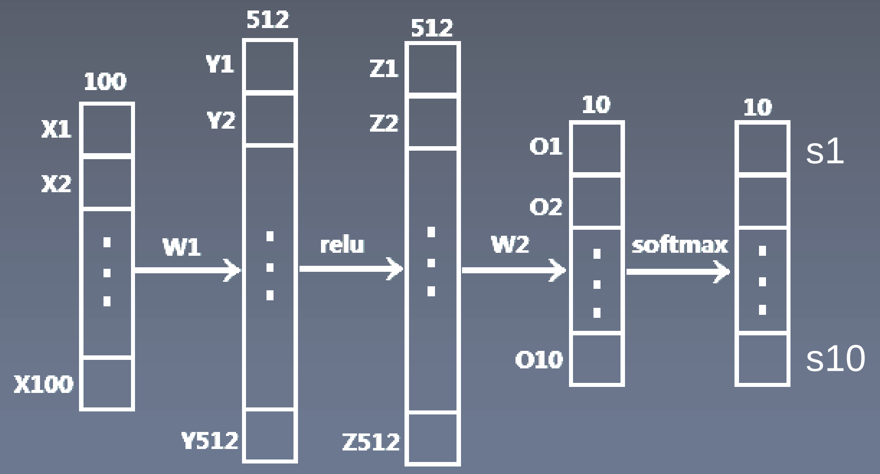
$$\frac{\partial J}{\partial o_m} = \frac{\partial\left(\ln \sum_{i=1}^{10} e^{o_i}\right)}{\partial o_m} - \frac{\partial(\ln e^{o_k})}{\partial o_m}$$

$$= \frac{e^{o_m}}{\sum_{i=1}^{10} e^{o_i}} - \delta(m = k)$$

$$= S_m - \delta(m = k)$$

$$\frac{\partial J}{\partial O} = S - \tilde{S} \quad 其中 \tilde{S} = (0, 0, \cdots, \overset{k}{1}, 0, \cdots, 0)$$

矩阵的求导在深度学习中的应用(中级)



$W1$: $100 * 512$ 的矩阵

$W2$: $512 * 10$ 的矩阵

$$(x_1, x_2, \cdots, x_{100})w_1 = (y_1, y_2, \cdots, y_{512})$$

$$(z_1, z_2, \cdots, z_{512})w_2 = (o_1, o_2, \cdots, o_{10})$$

$ZW = O$ 且 $\frac{\partial J}{\partial O} \overset{?}{\Rightarrow} \frac{\partial J}{\partial Z}, \frac{\partial J}{\partial W}$

为了简单：$Z = (z_1 \; z_2)$ $\quad W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$ $\quad O = (o_1 \; o_2 \; o_3)$

$$o_1 = w_{11}z_1 + w_{21}z_2$$
$$o_2 = w_{12}z_1 + w_{22}z_2$$
$$o_3 = w_{13}z_1 + w_{23}z_2$$

$$\frac{\partial J}{\partial z_1} = \frac{\partial J}{\partial o_1}\frac{\partial o_1}{\partial z_1} + \frac{\partial J}{\partial o_2}\frac{\partial o_2}{\partial z_1} + \frac{\partial J}{\partial o_3}\frac{\partial o_3}{\partial z_1} = \frac{\partial J}{\partial o_1}w_{11} + \frac{\partial J}{\partial o_2}w_{12} + \frac{\partial J}{\partial o_3}w_{13}$$

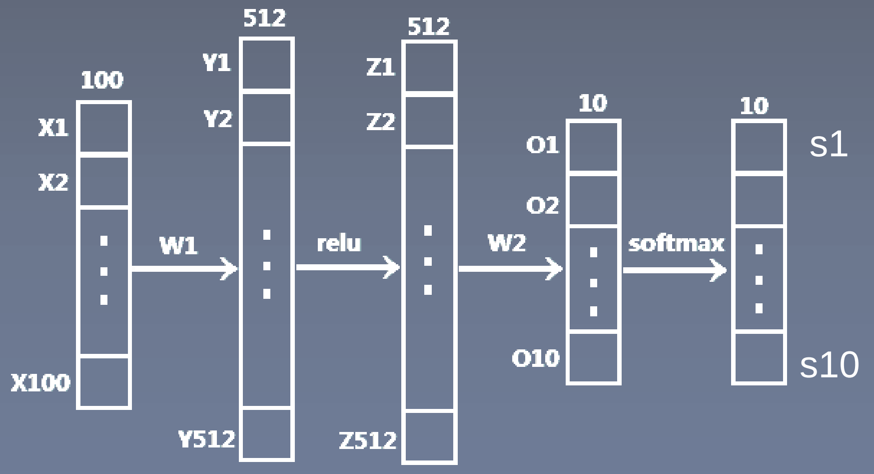
$$\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial o_1}w_{21} + \frac{\partial J}{\partial o_2}w_{22} + \frac{\partial J}{\partial o_3}w_{23}$$

$$\frac{\partial J}{\partial z} = \begin{pmatrix} \frac{\partial J}{\partial z_1} & \frac{\partial J}{\partial z_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial J}{\partial o_1} & \frac{\partial J}{\partial o_2} & \frac{\partial J}{\partial o_3} \end{pmatrix} \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{pmatrix} = \frac{\partial J}{\partial O} W^T$$

$$\frac{\partial J}{\partial W} = \begin{pmatrix} \frac{\partial J}{\partial o_1}z_1 & \frac{\partial J}{\partial o_2}z_1 & \frac{\partial J}{\partial o_3}z_1 \\ \frac{\partial J}{\partial o_1}z_2 & \frac{\partial J}{\partial o_2}z_2 & \frac{\partial J}{\partial o_3}z_2 \end{pmatrix}$$

$$= \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \begin{pmatrix} \frac{\partial J}{\partial o_1} & \frac{\partial J}{\partial o_2} & \frac{\partial J}{\partial o_3} \end{pmatrix} = Z^T \frac{\partial J}{\partial O}$$

矩阵的求导在深度学习中的应用(中级)



$W1$: $100 * 512$ 的矩阵
$W2$: $512 * 10$ 的矩阵

$$(x_1, x_2, \cdots, x_{100}) w_1 = (y_1, y_2, \cdots, y_{512})$$
$$(z_1, z_2, \cdots, z_{512}) w_2 = (o_1, o_2, \cdots, o_{10})$$

$$\frac{\partial J}{\partial w_2} = Z^T \frac{\partial J}{\partial O}$$
$$\frac{\partial J}{\partial Z} = \frac{\partial J}{\partial O} w_2^T$$

$$\frac{\partial J}{\partial w_1} = X^T \frac{\partial J}{\partial Y}$$
$$\frac{\partial J}{\partial X} = \frac{\partial J}{\partial Y} w_1^T$$

$$Z = relu(Y)$$
$$z_i = relu(y_i)$$

$$relu(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$$

$$\frac{\partial J}{\partial y_i} = \frac{\partial J}{\partial z_i} \frac{\partial z_i}{\partial y_i} = \frac{\partial J}{\partial z_i} \begin{cases} 1, y_i \geq 0 \\ 0, y_i < 0 \end{cases}$$

deepshare.net
深 度 之 眼

联系我们:

电话：18001992849

邮箱：service@deepshare.net

Q Q： 2677693114

公众号

客服微信