开闭

- 内点，内部interior

> 可以用任意范数定义一个球（范数等价性的体现）

- open开集合

> 不包含边界

- closed闭集合
- closure闭包

> 取int，再取补，保证闭包是闭的

The closure of a set C is defined as

$$\mathrm{cl}\, C = \mathbf{R}^n \setminus \mathrm{int}(\mathbf{R}^n \setminus C)$$

- boundary边界

$$\operatorname{bd} C = \operatorname{cl} C \setminus \operatorname{int} C$$

确界

- supremum上确界
- infimum下确界
- 关系

$$\inf C = -(\sup -C)$$

闭合函数

> 小于等于α的部分，对应的定义域是闭合的

- A function $f: \mathbf{R}^n \to \mathbf{R}$ is closed if, for each $\alpha \in \mathbf{R}$, the sublevel set

$$\{x \in \operatorname{dom} f \mid f(x) \leq \alpha\}$$

is closed. This is equivalent to

$\operatorname{epi} f = \{(x,t) \in \mathbf{R}^{n+1} | x \in \operatorname{dom} f, f(x) \leq t\}$ is closed

> 判定闭合函数等价方式：epi

导数

> 输入n维，输出m维，m*n维矩阵Df（x）
>
> 需要int，边界算不了

$$\lim_{z \in \operatorname{dom} f,\, z \neq x,\, z \to x} \frac{\|f(z) - f(x) - Df(x)(z-x)\|_2}{\|z-x\|_2} = 0$$

- 一阶近似--仿射函数affine（关于z的）

> 线性项+常数项，比如f（ax+B）
>
> 可以当作是f（z）的一阶近似，类似一阶泰勒展开

$$f(x) + Df(x)(z-x)$$

- 计算Df（x）

> 偏导计算导数矩阵

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, i = 1, \cdots, m, j = 1, \cdots, n$$

梯度gradient

> 当输出1维时，定义梯度为导数转置
>
> 用的就那么几个，有现成的，但需要自己算一遍

$$\nabla f(x) = Df(x)^{\mathsf{T}}$$

which is a column vector (in $\mathbf{R}^n$). Its components are the partial derivatives of f:

$$\nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, i = 1, \cdots, n$$

- 一阶近似--梯度版

$$f(x) + \nabla f(x)^{\mathsf{T}}(z - x)$$

链式法则chain

- Suppose $f: \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $x \in \text{int dom} f$ and $g: \mathbf{R}^m \to \mathbf{R}^p$ is differentiable at $f(x) \in \text{int dom}\, g$.

Define the composition $h: \mathbf{R}^n \to \mathbf{R}^p$ by $h(z) = g(f(z))$. Then h is differentiable at x, with derivate

$$Dh(x) = Dg(f(x))Df(x)$$

- 仿射特例

> ax+b这里是仿射函数

$$g(x) = f(Ax + b)$$

$$\nabla g(x) = A^{\mathsf{T}} \nabla f(Ax + b)$$

> f:Rn->R,g:R->R;a,b:Rn->Rn

```
gra(g(x))=aT*gra (f(ax+b))
```

```
g'(x)=bT*gra（f(ax+b))
```

二阶导数

> Rn->R
>
> hessian：二阶导组成矩阵

- Suppose $f: \mathbf{R}^n \to \mathbf{R}$. The second derivative or Hessian matrix of f at $x \in \operatorname{int} \operatorname{dom} f$, denoted $\nabla^2 f(x)$, is given by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i = 1, \cdots, n, j = 1, \cdots, n.$$

- 二阶近似

> 一阶后面再加一项

$$f(x) + \nabla f(x)^\mathsf{T}(z - x) + \frac{1}{2}(z - x)^\mathsf{T} \nabla^2 f(x)(z - x)$$

二阶导情况1

> g标量函数
>
> gra（f)是列向量，需要乘上自己转置变成矩阵

- Suppose $f: \mathbf{R}^n \to \mathbf{R}$, $g: \mathbf{R} \to \mathbf{R}$, and $h(x) = g(f(x))$

$$\nabla h(x) = g'\big(f(x)\big) \nabla f(x)$$

$$\nabla^2 h(x) = g'(f(x)) \nabla^2 f(x) + g''(f(x)) \nabla f(x) \nabla f(x)^\mathsf{T}$$

二阶导情况2

> 仿射作为输入得到g
>
> 两边是A，比较好记

Composition with affine function

$$g(x) = f(Ax + b)$$

$$\nabla g(x) = A^\top \nabla f(Ax + b)$$

$$\nabla^2 g(x) = A^\top \nabla^2 f(Ax + b) A$$

- 如果g本身是标量函数

$$g(t) = f(x + tv), \qquad x, v \in \mathbf{R}^n$$

$$g'(t) = v^\top \nabla f(x + tv)$$

$$g''(t) = v^\top \nabla^2 f(x + tv) v$$

常用导数梯度

1、二次函数

$$f(x) = \frac{1}{2} x^\top P x + q^\top x + r$$

$$\nabla f(x) = Px + q$$

$$\nabla^2 f(x) = P$$

2、logdet

> 输入是矩阵X，梯度也是矩阵，
> 需要两个下表索引，二阶导就需要4个索引,太复杂。不如直接写二阶近似

$$f(X) = \log\det X, \mathrm{dom}\, f = \mathbf{S}_{++}^n$$

$$\nabla f(X) = X^{-1}$$

$$f(X) + \mathrm{tr}\big(X^{-1}(Z - X)\big) - \frac{1}{2}\mathrm{tr}\big(X^{-1}(Z - X)X^{-1}(Z - X)\big)$$

3、例子：多维化归

- 题目

$$f(x) = \log \sum_{i=1}^{m} \exp(a_i^\top x + b_i)$$

> 合并之后，求导更容易，为了用链式法则

- 一阶导

$$f = g(Ax + b)$$

$$g(y) = \log \sum_{i=1}^{m} \exp(y_i)$$

$$\nabla g(y) = \frac{1}{\sum_{i=1}^{m} \exp y_i} \begin{bmatrix} \exp y_1 \\ \vdots \\ \exp y_m \end{bmatrix}$$

> 再用链式法则

$$\nabla f(x) = A^\top \nabla g(Ax + b) = \frac{1}{1^\top z} A^\top z$$

$$z = \begin{bmatrix} \exp a_1^\top x + b_1 \\ \vdots \\ \exp a_m^\top x + b_m \end{bmatrix}$$

> 1Tz：表示对z内求和

- 二阶导

> 先情况二，再情况一，用两次链式法则
> 求Hessian矩阵时，ij位置的元素即为对上述$\nabla f(x)$第i个元素关于x_j求导

$$\nabla^2 f(x) = A^\top \nabla g^2(Ax + b)A$$

> diag是对角矩阵
> det行列式

$$\nabla^2 g(y) = \mathrm{diag}(\nabla g(y)) - \nabla g(y) \nabla g(y)^\top$$

$$
\begin{aligned}
\nabla^2 f(x) &= A^\top \nabla g^2(Ax+b)A \\
&= A^\top \left(\frac{1}{1^\top z} \operatorname{diag}(z) - \frac{1}{(1^\top z)^2} z z^\top \right) A
\end{aligned}
$$

- $z_i = \exp(a_i^\top x + b_i), i = 1, \ldots, m$

4、例子2：

- 题目

$$
f(x) = \log\det(F_0 + x_1 F_1 + \cdots + x_n F_n)
$$

- 分解

$$
f(x) = g(F_0 + x_1 F_1 + \cdots + x_n F_n)
$$

$$
g(X) = \log\det X
$$

- 求偏导

```
利用g'(x)=bT*gra（f(ax+b)），x代表xi
```

> Fi为对内层xi求导结果，再对外面算梯度，
> 再两个矩阵算内积，即tr

$$
\frac{\partial f(x)}{\partial x_i} = \operatorname{tr}(F_i \nabla \log\det(F)) = \operatorname{tr}(F^{-1} F_i)
$$

> 再拼在一起

$$
\nabla f(x) = \begin{bmatrix} \operatorname{tr}(F^{-1} F_1) \\ \vdots \\ \operatorname{tr}(F^{-1} F_n) \end{bmatrix}
$$