数学

1.矩阵微积分

分子布局 分母布局的区别：一个标量关于一个向量的导数写成行向量还是列向量

(1) ① 标量关于向量的导数.

p维向量 $x \in R^p$, $y = f(x) = f(x_1, x_2 \cdots x_p) \in R$

$\frac{\partial y}{\partial x}$：分母布局 $\frac{\partial y}{\partial x} = [\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots \frac{\partial y}{\partial x_p}]^T \in R^{p\times 1}$ (列)

分子布局 $\frac{\partial y}{\partial x} = [\frac{\partial y}{\partial x_1} \cdots \frac{\partial y}{\partial x_p}] \in R^{1\times p}$ (行)

② 向量关于标量的导数.

$x \in R$, $y = f(x) \in R^q$, 则

分母布局 $\frac{\partial y}{\partial x} = [\frac{\partial y_1}{\partial x}, \frac{\partial y_2}{\partial x} \cdots \frac{\partial y_q}{\partial x}] \in R^{1\times q}$ (行)

分子布局 $\frac{\partial y}{\partial x} = [\frac{\partial y_1}{\partial x}; \cdots \frac{\partial y_q}{\partial x}] \in R^{q\times 1}$ (列)

③ 向量关于向量的导数

$x \in R^p$, $y = f(x) \in R^q$

分母布局 $\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_q}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_q}{\partial x_p} \end{bmatrix} \in R^{p\times q}$

(2) 导数法则 $x \in R^p$ $y = f(x) \in R^q$ $z = g(x) \in R^q$

① 加：$\frac{\partial (y+z)}{\partial x} = \frac{\partial y}{\partial x} + \frac{\partial z}{\partial x} \in R^{p\times q}$

② 乘：$\frac{\partial y^T z}{\partial x} = \frac{\partial y}{\partial x} z + \frac{\partial z}{\partial x} y \in R^p$

$\frac{\partial y^T A z}{\partial x} = \frac{\partial y}{\partial x} A z + \frac{\partial z}{\partial x} A y \in R^p$

y是标量 $\frac{\partial y z}{\partial x} = y\frac{\partial z}{\partial x} + \frac{\partial y}{\partial x} z^T \in R^{p\times q}$

(3) 链式法则：

(1) $x \in R$, $u = u(x) \in R^s$, $g = g(u) \in R^t$

$\frac{\partial g}{\partial x} = \frac{\partial u}{\partial x} \frac{\partial g}{\partial u} \in R^{1\times t}$

(2) $x \in R^p$, $y = g(x) \in R^s$ $z = f(y) \in R^t$.

$\frac{\partial z}{\partial x} = \frac{\partial y}{\partial x} \frac{\partial z}{\partial y} \in R^{1\times t}$

(3) $X \in R^{p\times q}$, $y = g(X) \in R^s$, $z = f(y) \in R$

$\frac{\partial z}{\partial X_{ij}} = \frac{\partial y}{\partial X_{ij}} \frac{\partial z}{\partial y} \in R$

2 逻辑斯谛函数

$logistic(x) = \frac{L}{1 + e^{-k(x-x_0)}}$

标准：$\frac{1}{1+e^{-x}} = \sigma(x)$

将实数空间映射到 (0,1) 区间

$\sigma'(x) = \sigma(x)(1-\sigma(x))$

K维向量 $\sigma'(x) = diag(\sigma(x) \odot (1-\sigma(x)))$

3 softmax 函数.

将多个标量映射到一个概率分布

对于K个标量 $x_1 \cdots x_k$.

$z_k = softmax(x_k) = \frac{exp(x_k)}{\sum_{i=1}^{k} exp(x_i)}$

输入K维向量 x, $\hat{z} = softmax(x)$

$= \frac{1}{\sum_{k=1}^{k} exp(x_k)} \begin{bmatrix} exp(x_1) \\ \vdots \\ exp(x_k) \end{bmatrix}$

$= \frac{exp(x)}{\sum_{k=1}^{k} exp(x_k)}$

$= \frac{exp(x)}{1_k^T exp(x)}$

softmax 函数的导数.

$\frac{\partial softmax(x)}{\partial x} = \frac{\partial \left(\frac{exp(x)}{1_k^T exp(x)}\right)}{\partial x}$

$= \frac{1}{1_k^T exp(x)} \frac{\partial exp(x)}{\partial x} + \frac{\partial \frac{1}{1_k^T exp(x)}}{\partial x} (exp(x))^T$

$= \frac{diag(exp(x))}{1_k^T exp(x)} - \left(\frac{1}{1_k^T exp(x)}\right)^2 \frac{\partial (1_k^T exp(x))}{\partial x} (exp(x))^T$

$= \frac{diag(exp(x))}{1_k^T exp(x)} - \frac{1}{(1_k^T exp(x))^2} exp(x)\, exp(x)^T$

$= diag\left(\frac{exp(x)}{1_k^T exp(x)}\right) - \frac{exp(x)}{1_k^T exp(x)} \cdot \frac{exp^T(x)}{1_k^T exp(x)}$

$= diag(softmax(x)) - softmax(x) \cdot softmax^T(x)$