机器学习数学基础

向量及向量的模

- 标量（Scalar）: 实数
- 向量（Vector）: n 个实数组成的有序数组，称为 n 维向量。如果没有特别说明，一个 n 维向量一般表示列向量，即大小为 $n \times 1$的矩阵
- 向量的模$\|a\|$表示向量a的大小

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \|a\| = \sqrt{\sum_{i=1}^{n} a_i^2}$$

向量的范数

- **范数(norm)：** 一种定义在赋范线性空间中函数，满足相应条件后的函数都可以被称为范数，是一个表示“长度”概念的函数，为向量空间内的所有向量赋予非零的正长度或大小。对于一个 n 维的向量a，其常见的范数有：

0- 范数：a中非零元素的个数

1- 范数：$|a|_1 = \sum_{i=1}^{n} |a_i|$

*2-*范数：$|a|_2 = \sqrt{\sum_{i=1}^{n} a_i^2} = \sqrt{a^T a}$

∞范数：$|a|_\infty = \max_i |a_i|$

-∞范数：$|a|_{-\infty} = \min_i |a_i|$

p-范数：$|a|_p = \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}}$

常见的向量及类型

- 全 **1** 向量指所有值为 1 的向量。用$\mathbf{1}_n$表示，n 表示向量的维数。

$$\mathbf{1}_n = [1,1,\dots,1]^T$$

$$b = [0,1,0\dots,0]^T$$

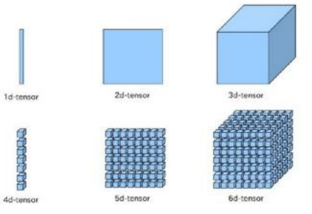
$$b = [0.3,-0.2,3.0,\dots,5]^T$$

矩阵和张量

- **矩阵（Matrix）**：一个大小为 $m \times n$ 的矩阵是一个由 m 行 n 列元素排列成的矩形阵列

$$A_{3\times3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

- **张量（Tensor）**：一个定义在一些向量空间和一些对偶空间的笛卡儿积上的多重线性映射，其坐标是|n|维空间内，有|n|个分量的一种量。
- 0阶张量：标量， 1阶张量：向量， 2阶张量：矩阵



5

5

矩阵的基本运算

- 如果 A 和 B 都为 $m \times n$ 的矩阵，则 A 和 B 的加减为

$$(A+B)_{ij} = A_{ij} + B_{ij}$$
$$(A-B)_{ij} = A_{ij} - B_{ij}$$

- 点乘 $(A \odot B)_{ij}$ 为　　$(A \odot B)_{ij} = A_{ij}B_{ij}$

- 一个标量 c 与矩阵 A 乘积为

$$(cA)_{ij} = cA_{ij}$$

- 若 A 是 $m \times p$ 矩阵和 B 是 $p \times n$ 矩阵，则乘积 AB 是一个 $m \times n$ 的矩阵

$$(AB)_{ij} = \sum_{k=1}^{p} A_{ik}B_{kj}$$

6

6

矩阵的基本运算

- 矩阵 $A = [a_{ij}]_{m\times n}$ 的转置（Transposition）是一个 $n \times m$ 的矩阵，记为 A^T

$$A_{3\times3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad A_{3\times3}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

- 矩阵的向量化：将矩阵表示为一个列向量。设 $A = [a_{ij}]_{m\times n}$，则

$$vec(A) = [a_{11}, a_{21}, \cdots a_{m1}, \cdots, a_{1n}, a_{2n}, \cdots, a_{mn}]^T$$

7

7

常见的矩阵

- 对称矩阵：满足 $A = A^T$。
- 对角矩阵（Diagonal Matrix）：一个主对角线之外的元素皆为 0 的矩阵。

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

可以记为$diag(a)$, a为一个 n 维向量，满足$A_{ii} = a_i$

- $n \times n$ 的对角矩阵 $A = diag(a)$和 n 维向量b的乘积为一个 n 维向量

$$Ab = diag(a)b$$
$$(Ab)_i = a_i b_i$$

8

常见的矩阵

- 单位矩阵是一种特殊的的对角矩阵，其主对角线元素为 1，其余元素为0。n 阶单位矩阵 I_n，是一个 $n \times n$ 的方形矩阵。可以记为

$$I_n = diag(1, \cdots, 1)$$

- 一个矩阵和单位矩阵的乘积等于其本身。

$$AI = IA = A$$

导数

- 对于定义域和值域都是实数域的函数 $y = f(x)$;。若 $f(x)$ 在点 x_0 的某个邻域 Δx 内，极限

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

存在，则称函数 $y = f(x)$ 在点 x_0 处可导，并导数为 $f'(x_0)$。若函数 $f(x)$ 在其定义域包含的某区间内每一个点都可导，那么也可以说函数 $f(x)$ 在这个区间内可导。这样，我们可以定义函数 $f'(x)$ 为函数 $f(x)$的**导函数**，通常也成为**导数**。

$$f'(x) \qquad \nabla_x f(x) \qquad \frac{\partial}{\partial x} f(x)$$

向量求导

- 向量对元素求导

对于一个元素$x \in R$，函数 $f(x) = [f_1(x) \cdots f_n(x)]^T \in R^n$是一个列向量，则 $f(x)$关于x的导数为

$$\frac{\partial}{\partial x} f(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x} \\ \vdots \\ \frac{\partial f_n(x)}{\partial x} \end{bmatrix}$$

- 元素对向量求导

对于一个 n 维列向量$x = [x_1, \cdots, x_n]^T \in R^n$，函数 $f(x) = f(x_1, \cdots, x_n) \in R$，则 $f(x)$关于x的导数为

$$\frac{\partial}{\partial x} f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

11

向量求导

向量对向量求导： 对于一个列向量$x \in R^m$，函数$f(x) \in R^n$是一个列向量，
则 $f(x)$关于 x的导数为

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

$$\frac{\partial}{\partial x} f(x)^T = \begin{bmatrix} \frac{\partial f(x)^T}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)^T}{\partial x_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_m} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{bmatrix} \qquad \frac{\partial}{\partial x^T} f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_m} \end{bmatrix}^T = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_m} \end{bmatrix}$$

12

矩阵求导

矩阵对元素求导： 对于一个元素$x \in R$，函数$f(x) \in R^{m\times n}$，则 $f(x)$关于 x的导数为

$$f(x)=\begin{bmatrix} f_{11}(x) & \cdots & f_{1n}(x) \\ \vdots & \ddots & \vdots \\ f_{m1}(x) & \cdots & f_{mn}(x) \end{bmatrix} \qquad x \qquad \frac{\partial}{\partial x}f(x)=\begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x} & \cdots & \frac{\partial f_{1n}(x)}{\partial x} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{m1}(x)}{\partial x} & \cdots & \frac{\partial f_{mn}(x)}{\partial x} \end{bmatrix}$$

- 元素对矩阵求导： 对于一个矩阵$x \in R^{m\times n}$，函数 $f(x) \in R$， 则 $f(x)$关于 x的导数为

$$f(x) \qquad x=\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \qquad \frac{\partial}{\partial x}f(x)=\begin{bmatrix} \frac{\partial f(x)}{\partial x_{11}} & \cdots & \frac{\partial f(x)}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x)}{\partial x_{m1}} & \cdots & \frac{\partial f(x)}{\partial x_{mn}} \end{bmatrix}$$

13

导数运算法则

- 加减法则：$y=f(x)\in R^n, z=g(x)\in R^n, x\in R^m$则

$$\frac{\partial}{\partial x}(y+z)=\frac{\partial y}{\partial x}+\frac{\partial z}{\partial x} \qquad \frac{\partial}{\partial x}(y-z)=\frac{\partial y}{\partial x}-\frac{\partial z}{\partial x}$$

- 乘法法则：$y=f(x)\in R^n, z=g(x)\in R^n, x\in R^m$

$$\frac{\partial y^T z}{\partial x}=\frac{\partial y^T}{\partial x}z+\frac{\partial z^T}{\partial x}y$$

- 乘法法则：$y=f(x)\in R^n, z=g(x)\in R^k, x\in R^m$，且$A\in R^{n\times k}$是与$x$无关的矩阵

$$\frac{\partial y^T A z}{\partial x}=\frac{\partial y^T}{\partial x}Az+\frac{\partial z^T}{\partial x}A^T y$$

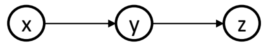
14

14

导数法则

- $y = f(x) \in R^n, z = g(y) \in R^k, x \in R^m$

$$\frac{\partial z}{\partial x} = \frac{\partial y}{\partial x}\frac{\partial z}{\partial y}$$



- 若 x 为一个矩阵，$y = f(x)$是一个矩阵， $z = g(y)$是元素

$$\frac{\partial z}{\partial x_{ij}} = tr\left(\left(\frac{\partial z}{\partial y}\right)^T \frac{\partial y}{\partial x_{ij}}\right)$$

- 若 x 为一个矩阵， $y = f(x)$是一个列向量， $z = g(y)$是元素

$$\frac{\partial z}{\partial x_{ij}} = \left(\frac{\partial z}{\partial y}\right)^T \frac{\partial y}{\partial x_{ij}}$$

向量求导

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

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

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

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

http://files.cnblogs.com/files/leoleo/matrix_rules.pdf

常用函数及其导数

指示函数

指示函数 $I(x = c)$ 为

$$I(x=c)==\begin{cases}1 & \text{if} \quad x=c\\ 0 & \text{else} \quad 0\end{cases}$$

指示函数 $I(x = c)$ 除了在 c 外，其导数为 0。

多项式函数

如果 $f(x) = x^r$，其中 r 是非零实数，那么导数

$$\frac{\partial x^r}{\partial x} = r x^{r-1}$$

当 $r = 0$ 时，常函数的导数是 0。

指数函数 $\exp(x) = e^x$

$$\frac{\partial \exp(x)}{\partial x} = \exp(x)$$

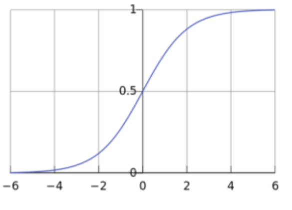
对数函数 $\log(x)$

$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$$

Logistic 函数

logistic 函数经常用来将一个实数空间的数映射到 (0,1) 区间，记为 $\sigma(x)$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



其导数为

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

当输入为 K 维向量 $x = [x_1, \cdots, x_K]^T$时，其导数为

$$\sigma'(x) = \mathrm{diag}(\sigma(x) \odot (1 - \sigma(x)))$$

Softmax 函数

softmax 函数是将多个标量映射为一个概率分布。
对于 K 个标量 $x_1, \cdots, x_K$，softmax 函数定义为

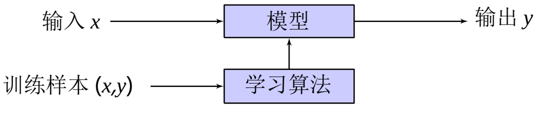
$$z_k = \mathrm{soft\,max}\left(x_k\right) = \frac{\exp\left(x_k\right)}{\sum_{i=1}^{K} \exp\left(x_i\right)}$$

19

19

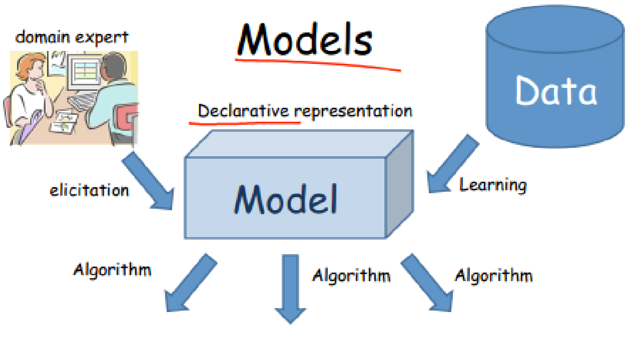
机器学习

- 机器学习主要是研究如何使计算机从给定的数据中学习规律，即从观测数据（样本）中寻找规律，并利用学习到的规律（模型）对未知或无法观测的数据进行预测。目前，主流的机器学习算法是基于统计的方法，也叫统计机器学习。



20

20


domain expert
Models
Data
Declarative representation
elicitation
Model
Learning
Algorithm
Algorithm
Algorithm