

AI基础：简易数学入门

原创：机器学习初学者 机器学习初学者 前天

0. 导语

学习AI建议掌握的数学资料：

数学分析（微积分），线性代数，概率论，统计，应用统计，数值分析，常微分方程，偏微分方程，数值偏微分方程，运筹学，离散数学，随机过程，随机偏微分方程，抽象代数，实变函数，泛函分析，复变函数，数学建模，拓扑，微分几何，渐近分析.....

可以劝退了.....

其实，绝大部分AI爱好者，对数学的要求没有那么高，只需要学习高等数学、线性代数、概率论与数理统计三门课，如果时间还不够，那看下本文我整理的必须掌握的部分。[看不懂公式的时候，可以查下本文，大部分能找到是什么意思。](#)

我最近在编写AI基础系列，数学是所有的基础。可以说，数学基础是机器学习从业人员的天花板。博士的代码能力，不一定比硕士强，但数学基础，往往要比硕士扎实很多。为什么机器学习从业人员学历越高，往往工资越高，通常和掌握的基础知识正相关。有时间，一定要认真打基础！！（黄海广）

目前已经发布：

AI 基础：Python开发环境设置和小技巧

AI 基础：Python 简易入门

AI 基础：Numpy 简易入门

AI 基础：Pandas 简易入门

AI 基础：Scipy(科学计算库) 简易入门

AI基础：数据可视化简易入门（matplotlib和seaborn）

AI基础：特征工程-类别特征

AI基础：特征工程-数字特征处理

AI基础：特征工程-文本特征处理

AI基础：词嵌入基础和Word2Vec

AI基础：图解Transformer

AI基础：一文看懂BERT

后续持续更新

本文节选自我的github里的内容，基本满足要求了，如果完整资料也可以在github下载：

<https://github.com/fengdu78/Data-Science-Notes>

高等数学

必须掌握导数和微分的概念

导数和微分的概念

$$f'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \quad (1)$$

或者：

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \quad (2)$$

四则运算法则

设函数 $u = u(x)$ ， $v = v(x)$ 在点 x 可导则：

$$(1) (u \pm v)' = u' \pm v'$$

$$d(u \pm v) = du \pm dv$$

$$(2) (uv)' = uv' + vu'$$

$$d(uv) = u dv + v du$$

基本导数与微分表

$$(1) y = c \text{ (常数)}$$

$$y' = 0$$

$$dy = 0$$

$$(2) y = x^\alpha (\alpha \text{为实数})$$

$$y' = \alpha x^{\alpha-1}$$

$$dy = \alpha x^{\alpha-1} dx$$

$$(3) y = a^x$$

$$y' = a^x \ln a$$

$$dy = a^x \ln a dx$$

$$\text{特例: } (e^x)' = e^x$$

$$d(e^x) = e^x dx$$

$$(4) y = \log_a x$$

$$y' = \frac{1}{x \ln a}$$

$$dy = \frac{1}{x \ln a} dx$$

$$\text{特例: } y = \ln x$$

$$(\ln x)' = \frac{1}{x}$$

$$d(\ln x) = \frac{1}{x} dx$$

$$(5) \left(\frac{u}{v}\right)' = \frac{vu' - uv'}{v^2} (v \neq 0)$$

$$d\left(\frac{u}{v}\right) = \frac{vdu - u dv}{v^2}$$

复合函数，反函数，隐函数以及参数方程所确定的函数的微分法

(1) 反函数的运算法则:

设 $y = f(x)$ 在点 x 的某邻域内单调连续，在点 x 处可导且 $f'(x) \neq 0$ ，则其反函数在点 x 所对应的 y 处可导，并且有 $\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}$

(2) 复合函数的运算法则:

若 $\mu = \varphi(x)$ 在点 x 可导,而 $y = f(\mu)$ 在对应点 $\mu(\mu = \varphi(x))$ 可导,则复合函数 $y = f(\varphi(x))$ 在点 x 可导,且 $y' = f'(\mu) \cdot \varphi'(x)$

(3) 隐函数导数 $\frac{dy}{dx}$ 的求法一般有三种方法:

1) 方程两边对 x 求导, 要记住 y 是 x 的函数, 则 y 的函数是 x 的复合函数. 例如 $\frac{1}{y}$, y^2 , $\ln y$, e^y 等均是 x 的复合函数.

对 x 求导应按复合函数连锁法则做.

2) 公式法. 由 $F(x, y) = 0$ 知 $\frac{dy}{dx} = -\frac{F'_x(x, y)}{F'_y(x, y)}$, 其中, $F'_x(x, y)$, $F'_y(x, y)$ 分别表示 $F(x, y)$ 对 x 和 y 的偏导数

3) 利用微分形式不变性

泰勒公式

设函数 $f(x)$ 在点 x_0 处的某邻域内具有 $n+1$ 阶导数, 则对该邻域内异于 x_0 的任意点 x , 在 x_0 与 x 之间至少存在一个 ξ , 使得: $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x)$ 其中 $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$ 称为 $f(x)$ 在点 x_0 处的 n 阶泰勒余项。

令 $x_0 = 0$, 则 n 阶泰勒公式 $f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + R_n(x) \dots\dots(1)$

其中 $R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}$, ξ 在0与 x 之间. (1)式称为麦克劳林公式

常用五种函数在 $x_0 = 0$ 处的泰勒公式

$$(1) e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + \frac{x^{n+1}}{(n+1)!}e^\xi$$

$$\text{或} = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{n!}x^n + o(x^n)$$

$$(2) \sin x = x - \frac{1}{3!}x^3 + \cdots + \frac{x^n}{n!}\sin \frac{n\pi}{2} + \frac{x^{n+1}}{(n+1)!}\sin(\xi + \frac{n+1}{2}\pi)$$

$$\text{或} = x - \frac{1}{3!}x^3 + \cdots + \frac{x^n}{n!}\sin \frac{n\pi}{2} + o(x^n)$$

$$(3) \cos x = 1 - \frac{1}{2!}x^2 + \cdots + \frac{x^n}{n!}\cos \frac{n\pi}{2} + \frac{x^{n+1}}{(n+1)!}\cos(\xi + \frac{n+1}{2}\pi)$$

$$\text{或} = 1 - \frac{1}{2!}x^2 + \cdots + \frac{x^n}{n!}\cos \frac{n\pi}{2} + o(x^n)$$

$$(4) \ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n-1}\frac{x^n}{n} + \frac{(-1)^n x^{n+1}}{(n+1)(1+\xi)^{n+1}}$$

$$\text{或} = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + (-1)^{n-1}\frac{x^n}{n} + o(x^n)$$

$$(5) (1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \cdots + \frac{m(m-1)\cdots(m-n+1)}{n!}x^n + \frac{m(m-1)\cdots(m-n+1)}{(n+1)!}x^{n+1}(1+\xi)^{m-n-1}$$

$$\text{或 } (1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \cdots + \frac{m(m-1)\cdots(m-n+1)}{n!}x^n + o(x^n)$$

线性代数

矩阵

矩阵： $m \times n$ 个数 a_{ij} 排成 m 行 n 列的表格 $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ 称为矩阵，简记为 A ，或者

$(a_{ij})_{m \times n}$ 。若 $m = n$ ，则称 A 是 n 阶矩阵或 n 阶方阵。

矩阵的线性运算

1. 矩阵的加法

设 $A = (a_{ij})$, $B = (b_{ij})$ 是两个 $m \times n$ 矩阵，则 $m \times n$ 矩阵 $C = (c_{ij}) = a_{ij} + b_{ij}$ 称为矩阵 A 与 B 的和，记为 $A + B = C$ 。

2. 矩阵的数乘

设 $A = (a_{ij})$ 是 $m \times n$ 矩阵， k 是一个常数，则 $m \times n$ 矩阵 (ka_{ij}) 称为数 k 与矩阵 A 的数乘，记为 kA 。

3. 矩阵的乘法

设 $A = (a_{ij})$ 是 $m \times n$ 矩阵， $B = (b_{ij})$ 是 $n \times s$ 矩阵，那么 $m \times s$ 矩阵 $C = (c_{ij})$ ，其中 $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^n a_{ik}b_{kj}$ 称为 AB 的乘积，记为 $C = AB$ 。

4. A^T 、 A^{-1} 、 A^* 三者之间的关系

$$(1) (A^T)^T = A, (AB)^T = B^T A^T, (kA)^T = kA^T, (A \pm B)^T = A^T \pm B^T$$

$$(2) (A^{-1})^{-1} = A, (AB)^{-1} = B^{-1}A^{-1}, (kA)^{-1} = \frac{1}{k}A^{-1},$$

但 $(A \pm B)^{-1} = A^{-1} \pm B^{-1}$ 不一定成立。

$$(3) (A^*)^* = |A|^{n-2} A \quad (n \geq 3), (AB)^* = B^* A^*, (kA)^* = k^{n-1} A^* \quad (n \geq 2)$$

但 $(A \pm B)^* = A^* \pm B^*$ 不一定成立。

$$(4) (A^{-1})^T = (A^T)^{-1}, (A^{-1})^* = (AA^*)^{-1}, (A^*)^T = (A^T)^*$$

5.有关矩阵秩的结论

(1) 秩 $r(A)$ =行秩=列秩;

(2) $r(A_{m \times n}) \leq \min(m, n)$;

(3) $A \neq 0 \Rightarrow r(A) \geq 1$;

(4) $r(A \pm B) \leq r(A) + r(B)$;

(5) 初等变换不改变矩阵的秩

(6) $r(A) + r(B) - n \leq r(AB) \leq \min(r(A), r(B))$, 特别若 $AB = O$ 则: $r(A) + r(B) \leq n$

(7) 若 A^{-1} 存在 $\Rightarrow r(AB) = r(B)$; 若 B^{-1} 存在 $\Rightarrow r(AB) = r(A)$;

若 $r(A_{m \times n}) = n \Rightarrow r(AB) = r(B)$; 若 $r(A_{m \times s}) = n \Rightarrow r(AB) = r(A)$ 。

(8) $r(A_{m \times s}) = n \Leftrightarrow Ax = 0$ 只有零解

向量

1.有关向量组的线性表示

(1) $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性相关 \Leftrightarrow 至少有一个向量可以用其余向量线性表示。

(2) $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性无关, $\alpha_1, \alpha_2, \dots, \alpha_s, \beta$ 线性相关 $\Leftrightarrow \beta$ 可以由 $\alpha_1, \alpha_2, \dots, \alpha_s$ 唯一线性表示。

(3) β 可以由 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表示 $\Leftrightarrow r(\alpha_1, \alpha_2, \dots, \alpha_s) = r(\alpha_1, \alpha_2, \dots, \alpha_s, \beta)$ 。

2.有关向量组的线性相关性

(1) 部分相关, 整体相关; 整体无关, 部分无关。

(2) ① n 个 n 维向量 $\alpha_1, \alpha_2 \dots \alpha_n$ 线性无关 $\Leftrightarrow |[\alpha_1 \alpha_2 \dots \alpha_n]| \neq 0$, n 个 n 维向量 $\alpha_1, \alpha_2 \dots \alpha_n$ 线性相关 $\Leftrightarrow |[\alpha_1, \alpha_2, \dots, \alpha_n]| = 0$ 。

② $n + 1$ 个 n 维向量线性相关。

③ 若 $\alpha_1, \alpha_2 \dots \alpha_s$ 线性无关, 则添加分量后仍线性无关; 或一组向量线性相关, 去掉某些分量后仍线性相关。

3.有关向量组的线性表示

(1) $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性相关 \Leftrightarrow 至少有一个向量可以用其余向量线性表示。

(2) $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性无关, $\alpha_1, \alpha_2, \dots, \alpha_s, \beta$ 线性相关 $\Leftrightarrow \beta$ 可以由 $\alpha_1, \alpha_2, \dots, \alpha_s$ 唯一线性表示。

(3) β 可以由 $\alpha_1, \alpha_2, \dots, \alpha_s$ 线性表示 $\Leftrightarrow r(\alpha_1, \alpha_2, \dots, \alpha_s) = r(\alpha_1, \alpha_2, \dots, \alpha_s, \beta)$

4. 向量组的秩与矩阵的秩之间的关系

设 $r(A_{m \times n}) = r$, 则 A 的秩 $r(A)$ 与 A 的行列向量组的线性相关性关系为:

(1) 若 $r(A_{m \times n}) = r = m$, 则 A 的行向量组线性无关。

(2) 若 $r(A_{m \times n}) = r < m$, 则 A 的行向量组线性相关。

(3) 若 $r(A_{m \times n}) = r = n$, 则 A 的列向量组线性无关。

(4) 若 $r(A_{m \times n}) = r < n$, 则 A 的列向量组线性相关。

5. n 维向量空间的基变换公式及过渡矩阵

若 $\alpha_1, \alpha_2, \dots, \alpha_n$ 与 $\beta_1, \beta_2, \dots, \beta_n$ 是向量空间 V 的两组基, 则基变换公式为:

$$(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1, \alpha_2, \dots, \alpha_n) \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} = (\alpha_1, \alpha_2, \dots, \alpha_n) C$$

其中 C 是可逆矩阵, 称为由基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵。

6. 坐标变换公式

若向量 γ 在基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 与基 $\beta_1, \beta_2, \dots, \beta_n$ 的坐标分别是 $X = (x_1, x_2, \dots, x_n)^T$,

$Y = (y_1, y_2, \dots, y_n)^T$ 即: $\gamma = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = y_1\beta_1 + y_2\beta_2 + \cdots + y_n\beta_n$, 则向量坐标变换公式为 $X = CY$ 或 $Y = C^{-1}X$, 其中 C 是从基 $\alpha_1, \alpha_2, \dots, \alpha_n$ 到基 $\beta_1, \beta_2, \dots, \beta_n$ 的过渡矩阵。

7. 向量的内积

$$(\alpha, \beta) = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \alpha^T\beta = \beta^T\alpha$$

概率论和数理统计

1. 事件的关系与运算

- (1) 子事件： $A \subset B$ ，若 A 发生，则 B 发生。
- (2) 相等事件： $A = B$ ，即 $A \subset B$ ，且 $B \subset A$ 。
- (3) 和事件： $A \cup B$ （或 $A + B$ ）， A 与 B 中至少有一个发生。
- (4) 差事件： $A - B$ ， A 发生但 B 不发生。
- (5) 积事件： $A \cap B$ （或 AB ）， A 与 B 同时发生。
- (6) 互斥事件（互不相容）： $A \cap B = \emptyset$ 。
- (7) 互逆事件（对立事件）： $A \cap B = \emptyset, A \cup B = \Omega, A = \bar{B}, B = \bar{A}$

2.运算律

- (1) 交换律： $A \cup B = B \cup A, A \cap B = B \cap A$
- (2) 结合律： $(A \cup B) \cup C = A \cup (B \cup C)$
- (3) 分配律： $(A \cap B) \cap C = A \cap (B \cap C)$

3.德·摩根律

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

4.完全事件组

$A_1 A_2 \cdots A_n$ 两两互斥，且和事件为必然事件，即 $A_i \cap A_j = \emptyset, i \neq j, \bigcup_{i=1}^n A_i = \Omega$

5.概率的基本公式

- (1) 条件概率：

$P(B|A) = \frac{P(AB)}{P(A)}$ ，表示 A 发生的条件下， B 发生的概率。

- (2) 全概率公式：

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i), B_i B_j = \emptyset, i \neq j, \bigcup_{i=1}^n B_i = \Omega$$

- (3) Bayes公式：

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}, j = 1, 2, \cdots, n \text{ 注：上述公式中事件 } B_i \text{ 的个数可为可列个。}$$

- (4) 乘法公式：

$$P(A_1 A_2) = P(A_1)P(A_2|A_1) = P(A_2)P(A_1|A_2)$$

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$$

6.事件的独立性

$$(1) A \text{与} B \text{相互独立} \Leftrightarrow P(AB) = P(A)P(B)$$

$$(2) A, B, C \text{两两独立}$$

$$\Leftrightarrow P(AB) = P(A)P(B); P(BC) = P(B)P(C); P(AC) = P(A)P(C);$$

$$(3) A, B, C \text{相互独立}$$

$$\Leftrightarrow P(AB) = P(A)P(B);$$

$$P(BC) = P(B)P(C); P(AC) = P(A)P(C); P(ABC) = P(A)P(B)P(C)$$

7.独立重复试验

将某试验独立重复 n 次，若每次实验中事件 A 发生的概率为 p ，则 n 次试验中 A 发生 k 次的概率为： $P(X = k) = C_n^k p^k (1 - p)^{n-k}$

8.重要公式与结论

$$(1) P(\bar{A}) = 1 - P(A)$$

$$(2) P(A \cup B) = P(A) + P(B) - P(AB)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$$

$$(3) P(A - B) = P(A) - P(AB)$$

$$(4) P(A\bar{B}) = P(A) - P(AB), P(A) = P(AB) + P(A\bar{B}),$$

$$P(A \cup B) = P(A) + P(\bar{A}B) = P(AB) + P(A\bar{B}) + P(\bar{A}B)$$

$$(5) \text{条件概率} P(\cdot|B) \text{满足概率的所有性质,}$$

例如：

$$P(\bar{A}_1|B) = 1 - P(A_1|B)$$

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 A_2|B)$$

$$P(A_1 A_2|B) = P(A_1|B)P(A_2|A_1 B)$$

$$(6) \text{若} A_1, A_2, \cdots, A_n \text{相互独立, 则} P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i), P\left(\bigcup_{i=1}^n A_i\right) = \prod_{i=1}^n (1 - P(\bar{A}_i))$$

$$(7) \text{互斥、互逆与独立性之间的关系:}$$

A 与 B 互逆 $\Rightarrow A$ 与 B 互斥，但反之不成立， A 与 B 互斥（或互逆）且均非零概率事件 $\Rightarrow A$ 与 B 不独立。

(8)若 $A_1, A_2, \dots, A_m, B_1, B_2, \dots, B_n$ 相互独立，则 $f(A_1, A_2, \dots, A_m)$ 与 $g(B_1, B_2, \dots, B_n)$ 也相互独立，其中 $f(\cdot), g(\cdot)$ 分别表示对相应事件做任意事件运算后所得的事件，另外，概率为1（或0）的事件与任何事件相互独立。

机器学习的常见推导

逻辑回归

逻辑回归代价函数：

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(h_{\theta}(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

$$\text{即： } J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

推导过程：

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))]$$

考虑：

$$h_{\theta}(x^{(i)}) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$

则：

$$\begin{aligned} & y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \\ &= y^{(i)} \log\left(\frac{1}{1 + e^{-\theta^T x^{(i)}}}\right) + (1 - y^{(i)}) \log\left(1 - \frac{1}{1 + e^{-\theta^T x^{(i)}}}\right) \\ &= -y^{(i)} \log(1 + e^{-\theta^T x^{(i)}}) - (1 - y^{(i)}) \log(1 + e^{\theta^T x^{(i)}}) \end{aligned}$$

所以：

$$\begin{aligned} \frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \left[-\frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(1 + e^{-\theta^T x^{(i)}}) - (1 - y^{(i)}) \log(1 + e^{\theta^T x^{(i)}})] \right] \\ &= -\frac{1}{m} \sum_{i=1}^m \left[-y^{(i)} \frac{-x_j^{(i)} e^{-\theta^T x^{(i)}}}{1 + e^{-\theta^T x^{(i)}}} - (1 - y^{(i)}) \frac{x_j^{(i)} e^{\theta^T x^{(i)}}}{1 + e^{\theta^T x^{(i)}}} \right] \end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{m} \sum_{i=1}^m y^{(i)} \frac{x_j^{(i)}}{1+e^{\theta^T x^{(i)}}} - (1 - y^{(i)}) \frac{x_j^{(i)} e^{\theta^T x^{(i)}}}{1+e^{\theta^T x^{(i)}}} \\
&= -\frac{1}{m} \sum_{i=1}^m \frac{y^{(i)} x_j^{(i)} - x_j^{(i)} e^{\theta^T x^{(i)}} + y^{(i)} x_j^{(i)} e^{\theta^T x^{(i)}}}{1+e^{\theta^T x^{(i)}}} \\
&= -\frac{1}{m} \sum_{i=1}^m \frac{y^{(i)} (1+e^{\theta^T x^{(i)}}) - e^{\theta^T x^{(i)}}}{1+e^{\theta^T x^{(i)}}} x_j^{(i)} \\
&= -\frac{1}{m} \sum_{i=1}^m \left(y^{(i)} - \frac{e^{\theta^T x^{(i)}}}{1+e^{\theta^T x^{(i)}}} \right) x_j^{(i)} \\
&= -\frac{1}{m} \sum_{i=1}^m \left(y^{(i)} - \frac{1}{1+e^{-\theta^T x^{(i)}}} \right) x_j^{(i)} \\
&= -\frac{1}{m} \sum_{i=1}^m [y^{(i)} - h_{\theta}(x^{(i)})] x_j^{(i)} \\
&= \frac{1}{m} \sum_{i=1}^m [h_{\theta}(x^{(i)}) - y^{(i)}] x_j^{(i)}
\end{aligned}$$

注：虽然得到的梯度下降算法表面上看上去与线性回归的梯度下降算法一样，但是这里的 $h_{\theta}(x) = g(\theta^T X)$ 与线性回归中不同，所以实际上是不一样的。另外，在运行梯度下降算法之前，进行特征缩放依旧是非常必要的。

最小二乘法

需要用到的性质：

- $\nabla_x b^T x = b$
- $\nabla_x x^T A x = 2Ax$ (如果 A 是对称阵)
- $\nabla_x^2 x^T A x = 2A$ (如果 A 是对称阵)

假设我们得到矩阵 $A \in \mathbb{R}^{m \times n}$ （为了简单起见，我们假设 A 是满秩）和向量 $b \in \mathbb{R}^m$ ，从而使 $b \notin \mathcal{R}(A)$ 。在这种情况下，我们将无法找到向量 $x \in \mathbb{R}^n$ ，由于 $Ax = b$ ，因此我们想要找到一个向量 x ，使得 Ax 尽可能接近 b ，用欧几里德范数的平方 $\|Ax - b\|_2^2$ 来衡量。

使用公式 $\|x\|^2 = x^T x$ ，我们可以得到：

$$\begin{aligned}
\|Ax - b\|_2^2 &= (Ax - b)^T (Ax - b) \\
&= x^T A^T A x - 2b^T A x + b^T b
\end{aligned}$$

根据 x 的梯度，并利用上面推导的性质：

$$\begin{aligned}
\nabla_x (x^T A^T A x - 2b^T A x + b^T b) &= \nabla_x x^T A^T A x - \nabla_x 2b^T A x + \nabla_x b^T b \\
&= 2A^T A x - 2A^T b
\end{aligned}$$

将最后一个表达式设置为零，然后解出 x ，得到了正规方程：

$$x = (A^T A)^{-1} A^T b$$

结语

本文节选自我的github里的内容，基本能满足要求了，如果需要完整资料可以在我的github下载：

<https://github.com/fengdu78/Data-Science-Notes>



备注：公众号菜单包含了整理了一本**AI小抄**，**非常适合在通勤路上用学习**。



往期回顾

- 那些年做的学术公益-你不是一个人在战斗
- 适合初学者入门人工智能的路线及资料下载
- 机器学习在线手册
- 深度学习在线手册

备注：加入本站微信群或者qq群，请回复“加群”

加入知识星球（4500+用户，ID：92416895），请回复“知识星球”

喜欢文章，点个在看👉

