机器学习

欧阳毅

浙江工商大学
管理工程与电子商务学院

2023

Contents I

考核要求

- 考试形式：闭卷
- 成绩计算方式：平时成绩 40% + 考试成绩 60%
- 总分中各部分所占比例：课后作业 20%，上机实验 10%，考勤 10%，期末考试 60%

参考书目

- 教材：1.
- 参考书目：
  - 1. 模式分类 [美]Richard O. Dua Peter E. Hart David G. Stork 著, 李宏东，姚天翔译，机械工业出版社
  - 2.
- 课件：oyydata.top

机器学习究竟在研究什么？

Data → Learning algorithm → Knowledge

设计和分析算法的目的是：

- 利用经验知识，在某些任务中，改进算法的效率

机器学习可以用于做什么？

- 文档分类

机器学习可以用于做什么？

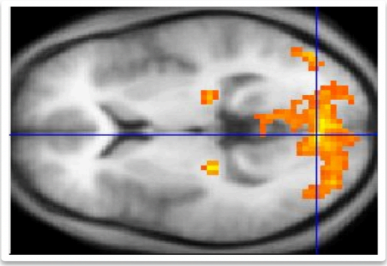
- 垃圾邮件过滤

机器学习可以用于做什么？

- 股市预测

机器学习可以用于做什么?

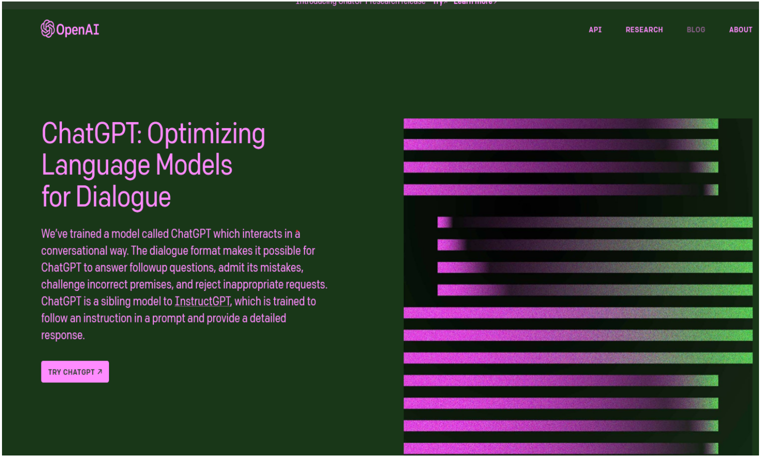
- 医学诊断



→ 核磁共振图像中有肿瘤分险

机器学习可以用于做什么？

- 自然语言处理 (ChatGPT)



机器学习可以用于做什么？

- 语音识别
- 机器视觉
- 机器人控制
- 医药分析
- 传感器网络
- 社交网络
- ……

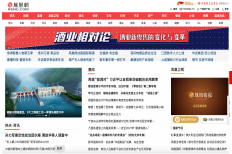
机器学习的分类

按学习任务可分为两大类:

- 监督方式学习
  - 分类
  - 回归
- 非监督方式学习
  - 密度估计
  - 聚类
  - 降维
- 半监督方式学习
- 增强学习

监督方式学习

Supervised Learning

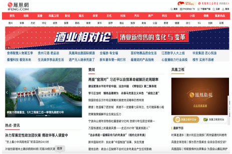
特征空间 $\mathcal{X}$	标签空间 $\mathcal{Y}$	
	体育 社会政治 科学...	离散标签 分类
	价格:26	连续标签 回归

任务：给定 $X \in \mathcal{X}$, 预测 $Y \in \mathcal{Y}$

非监督方式学习
Unsupervised Learning

无导师方式进行学习

特征空间 $\mathcal{X}$



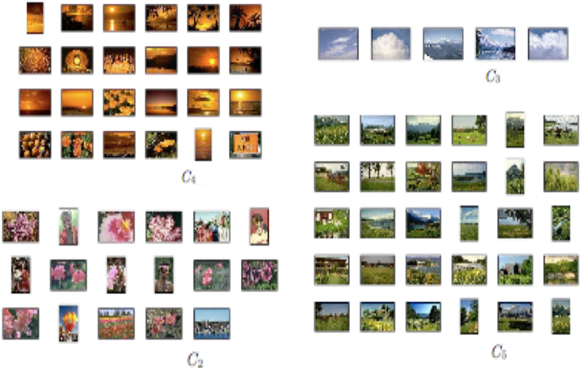
得到词分布

每个词出现的概率

任务：给定 $X \in \mathcal{X}$，学习 $f(X)$

非监督方式学习

Unsupervised Learning



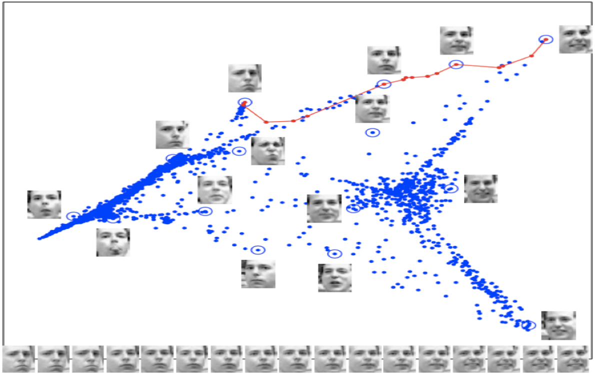
图像聚类

任务：给定 $X \in \mathcal{X}$, 学习 $f(X)$

非监督方式学习

Unsupervised Learning

降维处理



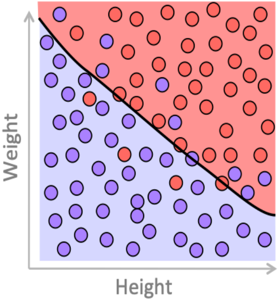
任务：给定 $X \in \mathcal{X}$，学习 $f(X)$

机器学习与统计学习的关系

目标： 构造一个预测规则 $f \colon \mathcal{X} \to \mathcal{Y}$

f 能比较好的反映测试数据样本点 $(X, Y) \sim P_{XY}$ 的分布情况

训练数据：



足球队员

其它人

目录

什么样的准则学习或选择最优的模型 I

监督学习问题用一个损失函数 (loss function) 来度量预测错误的程度, 损失函数是 f(X) 和 Y 的非负实值函数, 记作 L(Y,f(x)) 常用的损失函数有以下几种:

- (1)0-1 损失函数 (0-1 loss function)

$$L(Y,f(x))=\begin{cases}1 & , \quad Y\neq f(X)\\ 0 & , \quad Y=f(X)\end{cases} \tag{1}$$

- (2) 平方损失函数 (quadratic loss function)

$$L(Y,f(X))=(Y-f(X))^2$$

- (3) 绝对损失函数 (absolute loss function

$$L(Y,f(X))=|Y-f(X)|$$

什么样的准则学习或选择最优的模型 II

- (4) 对数损失函数或对数似然损失函数

$$L(Y, f(X)) = -logP(Y|X)$$

损失函数值越小, 模型就越好. 损失函数的期望:

$$R_{exp}(f) = E_P[L(Y, f(X))] = \int_{X\times Y} L(y, f(x))P(x, y)dxdy$$

这是理论上模型 $f(X)$ 关于联合分布 $P(X, Y)$ 的平均意义下的损失, 称为风险函数 (risk function) 或期望损失 (expected loss).

经验风险 (empirical risk)

- 模型 f(x) 关于训练数据集的平均损失称为经验风险 (empirical risk) 或经验损失 (empirical loss), 记作:

$$R_{emp}(f) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i))$$

- 期望风险 $R_{exp}(f)$ 是模型关于联合分布的期望损失
- 经验风险 $R_{emp}(f)$ 是模型关于训练样本集的平均损失

根据大数定律, 当样本容量 N 趋于无穷时, 经验风险 $R_{emp}(f)$ 趋于期望风险 $R_{exp}(f)$.

经验风险 (empirical risk)

但是, 由于现实中训练样本数目有限, 甚至很小, 所以用经验风险估计期望风险常常并不理想, 要对经验风险进行一定的矫正. 这就关系到监督学习的两个基本策略:

- 经验风险最小化
- 结构风险最小化

经验风险最小化和结构风险最小化

- 经验风险最小化 (empirical risk minimization,ERM) 的策略认为, 经验风险最小的模型是最优的模型.

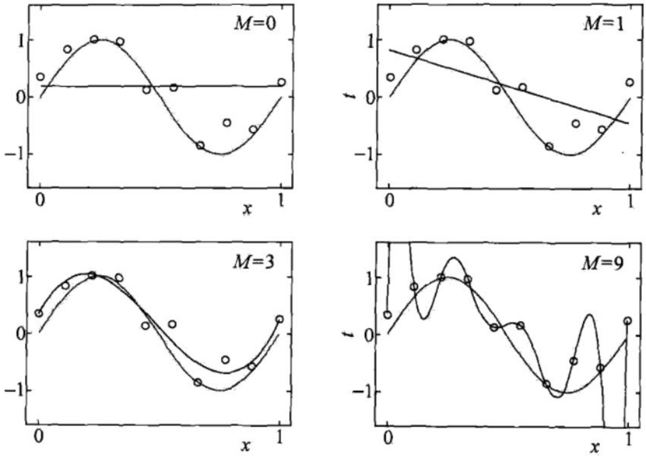
$$\min_{f\in F}\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$$

当样本容量很小时, 经验风险最小化学习的效果就未必很好, “过拟合 (over-fitting)”现象

过拟合 (over-fitting)

图: M 次多项式函数拟合问题的例子

$f_M(x,w) = w_0 + w_1x + w_2x^2 + ... + w_Mx^M$



结构风险最小化

- 结构风险最小化 (structural risk minimization,SRM)
  在经验风险上加上表示模型复杂度的正则化项 (regularizer) 或罚项.

$$R_{srm}(f) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i)) + \lambda J(f)$$

其中 J(f) 为模型的复杂度, 模型 f 越复杂, 复杂度 J 就越大; 反之, 模型 f 越简单, 复杂度 J(f) 就越小. 也就是说, 复杂度表示了对复杂模型的惩罚.

$$R_{srm}(f) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i)) + \lambda \sum_{i=1}^{m} \theta_i^2$$

目录

交叉验证 Cross Validation

- 交叉验证的目的是为了得到可靠稳定的模型
- 交叉验证 (cross-validation) 主要用于模型训练或建模应用中，如分类预测, 回归建模等。
- k 折交叉

交叉验证


Training Set
Training folds
Test fold
第 1 次迭代 → E_1
第 2 次迭代 → E_2
第 3 次迭代 → E_3
……
第 9 次迭代 → E_9
$E=\frac{1}{9}\sum_{i=1}^{9}E_i$

交叉验证

Step1: 将学习样本空间分为大小相等的 k 份
Step2: 选出第 i 份训练集作为测试集
Step3:
for j in (k-1 训练集) **do**
  训练第 i 个训练集，得到一个分类模型, 计算并保存模型评估指标
end for
Step4: 计算模型的平均性能
Step5: 用这 k 个模型在最终验证集的分类准确率平均值作为此 k-cv 下分类器的性能指标.

目录

评价指标

- 你训练了两个分类模型 $h_\theta(x), f_w(x)$ 来预测一个病人是否得了癌症

$$h_\theta(x) = \begin{cases} 1 & if \quad cancer, \\ 0 & \quad otherwise \end{cases}$$

$$f_w(x) = \begin{cases} 1 & if \quad cancer, \\ 0 & \quad otherwise \end{cases}$$

  你发现，你的分类器在测试集上的错误率是 1%（99% 的病人都正确诊断了）。只有 0.5% 的病人患有癌症。
- 如何判断哪个分类器更好?
- 在各类数据量不均衡时，分类准确率不是好的评价指标。

评价指标

处理不均衡数据的对策：建立新的评价指标。

- 混淆矩阵
- TP + TN + FP + TN = 样本总数

		真实的类别 1	0
预测	1	True Positive(TP)	False Positive(FP)
	0	False Negative (FN)	True Negative (TN)

评价指标

准确率 (Accuracy)：分类模型所有判断正确的结果占总观测值的比重

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{正确预测的样本}}{\text{样本总数}}$$

查准率 (Precision)：所有预测为患癌症的病人，有多少比例真正患了癌症?

$$Precision = \frac{TP}{TP + FP} = \frac{\text{正确预测为正的样本}}{\text{所有预测为正的样本}}$$

召回率 (Recall)：所有实际患癌症的病人中，有多少比例被检测出来了?

$$Recall = \frac{TP}{TP + FN} = \frac{\text{正确预测为正的样本}}{\text{所有真实为正的样本}}$$

评价指标

例

若混淆矩阵如下，请给出 Acc，Precision 和 Recall

	1	0
1	100	20
0	80	300

评价指标

例

若样本的预测标签和真实标签如下：请给出 Acc，Precision 和 Recall

```
预测标签    真实标签
[ 1           1],
[ 1           0],
[ 1,          1],
[ 0,          1],
[ 1           0],
[ 1,          1],
[ 0,          0],
[ 1           1],
[ 1,          0],
[ 0,          0]
```

评价指标

如果你的分类模型 $0 \leq h_\theta(x) \leq 1$ 已经确定了，提高一个值必然降低另一个值. 但一般情况下，Precision 高，Recall 就低，Recall 高，Precision 就低。

$$\begin{cases} \text{预测 } 1 & if \quad h_\theta(x) \geq 0.5 \\ \text{预测 } 0 & if \quad h_\theta(x) < 0.5 \end{cases}$$

- 如果我们想要只有在非常确信的时候预测 y=1 (提高查准率，降低误检率)，则提高阈值。
  高 precision，低 recall。
- 如果我们想要尽可能的不要漏掉潜在的癌症患者 (提高召回率，降低漏检率)，则降低阈值。
  低 precision，高 recall。

F-score

很多时候我们需要综合权衡这 2 个指标，这就引出了一个新的指标 F-score

$$F - Score = (1+\beta^2) \times \frac{Precision \times Recall}{\beta^2 \times Precision + Recall}$$

若混淆矩阵如下，请给出 F-score

	1	0
1	100	20
0	80	300

- Acc=100/500=20%
- Precision=100/120
- $F1 - score = 2 * \frac{20/120}{1\times100/120+100/180} =$
- $F2 - score = 5 * \frac{20/120}{4\times100/120+100/180} =$

当 $\beta = 1$ 时，称为 F1-score，这时，精确率和召回率都很重要。
当我们认为精确率更重要，调整 β 的值小于 1，
如果我们认为召回率更重要，调整 β 的值大于 1

目录

线性空间 I

1 设 E 是一个集合，在此集合上定义加法和数乘运算，并且运算是封闭的，这样形式的 E 称为线性空间，其中 E 的元素 x 称为向量

$$x = (x_1, x_2, ..., x_n)^T$$

2 线性组合: 设 $a_i \in E$ 中 m 个向量

$$x = \sum_{i=1}^{m} \lambda_i a_i$$

3 线性相关：如果存在 $\lambda_i \neq 0$(即 λ_i 不全为 0)，使得

$$\sum_{i=1}^{m} \lambda_i a_i = 0$$

反之则称 $a_1, a_2, ..., a_m$ 线性无关

线性空间 II

4 向量 x 与 y 的内积定义为:

$$< x, y >= \sum_{i=1}^{n} x_i y_i = x^T y$$

内积具有如下性质:

- $< x, x >\geq 0$, 且 $< x, x >= 0$ 的充分必要条件是 x=0
- $< x, y >=< y, x >$
- $< \lambda x + \mu y, z >= \lambda < x, z > +\mu < y, z >$

5 L_2 范数定义为:

$$||x|| = (\sum_{i=1}^{n} x_i^2)^{1/2} = (< x, x >)^{1/2}$$

范数具有如下性质:

- $||x|| \geq 0$, 且 $||x|| = 0$ 的充分必要条件是 x=0

线性空间 III

- $||\lambda x|| = |\lambda|||x||$
- $||x + y|| \leq ||x|| + ||y||$

6 若 x,y 满足:

$$< x, y >= 0$$

则称 x 与 y 正交

目录

矩阵和向量的求导法则 I

1 行向量对元素求导

设 $y^T = [y_1, ..., y_n]$ 是 n 维行向量，x 是元素，则

$$\frac{\partial y^T}{\partial x} = [\frac{\partial y_1}{\partial x}, ..., \frac{\partial y_n}{\partial x}]$$

2 列向量对元素求导

设 $y = \begin{bmatrix} y_1 \\ . \\ y_n \end{bmatrix}$ 是 n 维列向量，x 是元素，则

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ . \\ \frac{\partial y_n}{\partial x} \end{bmatrix}$$

矩阵和向量的求导法则 II

3 矩阵对元素求导

设 $Y = \begin{bmatrix} y_{11} \cdots y_{1n} \\ \cdots \\ y_{m1} \cdots y_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，x 是元素，则

$$\frac{\partial Y}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} \cdots \frac{\partial y_{1n}}{\partial x} \\ \cdots \\ \frac{\partial y_{m1}}{\partial x} \cdots \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$$

4 元素对行向量求导

设 y 是元素，$x^T = [x_1, ..., x_n]$ 是 n 维行向量则

$$\frac{\partial y}{\partial x^T} = [\frac{\partial y}{\partial x_1}, ..., \frac{\partial y}{\partial x_n}]$$

矩阵和向量的求导法则 III

5 元素对列向量求导

设 y 是元素,$x = \begin{bmatrix} x_1 \\ . \\ x_n \end{bmatrix}$ 是 n 维列向量，则

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ . \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$

6 元素对矩阵求导

设 y 是元素,$X = \begin{bmatrix} x_{11} ... x_{1n} \\ ... \\ x_{m1} ... x_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，则

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} \cdots \frac{\partial y}{\partial x_{1n}} \\ \cdots \\ \frac{\partial y}{\partial x_{m1}} \cdots \frac{\partial y}{\partial x_{mn}} \end{bmatrix}$$

矩阵和向量的求导法则 IV

7 行向量对列向量求导

设 $y^T=[y_1,...,y_n]$ 是 n 维行向量，x 是 $x=\begin{bmatrix} x_1 \\ . \\ x_m \end{bmatrix}$ m 维列向量，则

$$\frac{\partial y^T}{\partial x}=\begin{bmatrix} \frac{\partial y_1}{\partial x_1} \cdots \frac{\partial y_n}{\partial x_1} \\ \cdots \\ \frac{\partial y_n}{\partial x_m} \cdots \frac{\partial y_n}{\partial x_m} \end{bmatrix}$$

8 列向量对行向量求导

设 $y=\begin{bmatrix} y_1 \\ . \\ y_n \end{bmatrix}$ 是 m 维列向量,$x^T=[x_1,...,x_m]$ 是 n 维行向量，则

$$\frac{\partial y}{\partial x^T}=\begin{bmatrix} \frac{\partial y_1}{\partial x_1} \cdots \frac{\partial y_1}{\partial x_m} \\ \cdots \\ \frac{\partial y_n}{\partial x_1} \cdots \frac{\partial y_n}{\partial x_m} \end{bmatrix}$$

矩阵和向量的求导法则 V

9 行向量对行向量求导

设 $y^T = [y_1, ..., y_n]$ 是 n 维行向量，$x^T = [x_1, ..., x_m]$ 是 m 维行向量，则

$$\frac{\partial y^T}{\partial x^T} = [\frac{\partial y^T}{\partial x_1}, ..., \frac{\partial y^T}{\partial x_m}]$$

10 列向量对列向量求导

设 $y = \begin{bmatrix} y_1 \\ . \\ y_n \end{bmatrix}$ 是 n 维列向量，$x = \begin{bmatrix} x_1 \\ . \\ x_m \end{bmatrix}$ 是 m 维列向量，则

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ . \\ \frac{\partial y_n}{\partial x} \end{bmatrix}$$

矩阵和向量的求导法则 VI

11 矩阵对行向量求导

设 $Y = \begin{bmatrix} y_{11} \cdots y_{1n} \\ \cdots \\ y_{m1} \cdots y_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，$x^T = [x_1, ..., x_q]$ 是 q 维行向量，则

$$\frac{\partial Y}{\partial x^T} = \left[\frac{\partial Y}{\partial x_1} \cdots \frac{\partial Y}{\partial x_q} \right]$$

12 矩阵对列向量求导

设 $Y = \begin{bmatrix} y_{11} \cdots y_{1n} \\ \cdots \\ y_{m1} \cdots y_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，$x = \begin{bmatrix} x_1 \\ . \\ x_q \end{bmatrix}$ 是 q 维列向量，则

$$\frac{\partial Y}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} \cdots \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{m1}}{\partial x} \cdots \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$$

矩阵和向量的求导法则 VII

13 行向量对矩阵求导

设 $y^T=[y_1,...,y_q]$ 是 q 维行向量,$X=\begin{bmatrix} x_{11}...x_{1n} \\ ... \\ x_{m1}...x_{mn} \end{bmatrix}$ 是 $m\times n$ 矩阵，则

$$\frac{\partial y^T}{\partial X}=\begin{bmatrix} \frac{\partial y^T}{\partial x_{11}}\cdots\frac{\partial y^T}{\partial x_{1n}} \\ . \\ \frac{\partial y^T}{\partial x_{m1}}\cdots\frac{\partial y^T}{\partial x_{mn}} \end{bmatrix}$$

矩阵和向量的求导法则 VIII

14 列向量对矩阵求导

设 $y = \begin{bmatrix} y_1 \\ . \\ y_q \end{bmatrix}$ q 维列向量, 是 q 维行向量,$X = \begin{bmatrix} x_{11} \ldots x_{1n} \\ \ldots \\ x_{m1} \ldots x_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，则

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y_1}{\partial X} \\ . \\ \frac{\partial y_q}{\partial X} \end{bmatrix}$$

矩阵和向量的求导法则 IX

15 矩阵对矩阵求导

设 $Y = \begin{bmatrix} y_{11} \cdots y_{1n} \\ \cdots\cdot \\ y_{m1} \cdots y_{mn} \end{bmatrix} = \begin{bmatrix} y_1^T \\ . \\ y_m^T \end{bmatrix}$ 是 $m \times n$ 维矩阵,$X = \begin{bmatrix} x_{11} \ldots x_{1q} \\ \ldots \\ x_{p1} \ldots x_{pq} \end{bmatrix} = [x_1, \ldots, x_q]$ 是 $p \times q$ 矩阵，则

$$\frac{\partial Y}{\partial X} = \left[\frac{\partial Y}{\partial x_1} \cdots \frac{\partial Y}{\partial x_q} \right] = \begin{bmatrix} \frac{\partial y_1^T}{\partial x_1} \cdots \frac{\partial y_1^T}{\partial x_q} \\ \cdots \\ \frac{\partial y_m^T}{\partial x_1} \cdots \frac{\partial y_m^T}{\partial x_q} \end{bmatrix}$$

练习

- 若 X 是矩阵，θ 是向量，求:

$$\frac{\partial X\theta}{\partial \theta} = X$$

向量对向量求导，就是分子向量的每个分量对分母向量的每个分量求导，(1) 以（分母）作为主序

$$\frac{\partial X\theta}{\partial \theta} = \begin{bmatrix} \frac{\sum_{i=1}^{k} x_{1i}\theta_i}{\partial \theta_1} & \frac{\sum_{i=1}^{k} x_{1i}\theta_i}{\partial \theta_2} \dots \frac{\sum_{i=1}^{k} x_{1i}\theta_i}{\partial \theta_k} \\ & \dots \\ \frac{\sum_{i=1}^{k} x_{Ni}\theta_i}{\partial \theta_1} & \frac{\sum_{i=1}^{k} x_{Ni}\theta_i}{\partial \theta_2} \dots \frac{\sum_{i=1}^{k} x_{Ni}\theta_i}{\partial \theta_k} \end{bmatrix} \quad (2)$$

$$= \begin{bmatrix} x_{11} & x_{12} \dots x_{1k} \\ & \dots \\ x_{N1} & x_{N2} \dots x_{Nk} \end{bmatrix} \quad (3)$$

$$= X \quad (4)$$

练习

- 若 X 是矩阵，θ 是向量，求：

$$\frac{\partial X\theta}{\partial\theta} = X^T$$

向量对向量求导，就是分子向量的每个分量对分母向量的每个分量求导，(2) 以（分子）作为主序

$$\frac{\partial X\theta}{\partial\theta} = \begin{bmatrix} \frac{\sum_{i=1}^{k} x_{1i}\theta_i}{\partial\theta_1} & \frac{\sum_{i=1}^{k} x_{2i}\theta_i}{\partial\theta_1} \dots \frac{\sum_{i=1}^{k} x_{Ni}\theta_i}{\partial\theta_1} \\ & \dots \\ \frac{\sum_{i=1}^{k} x_{1i}\theta_i}{\partial\theta_k} & \frac{\sum_{i=1}^{k} x_{2i}\theta_i}{\partial\theta_k} \dots \frac{\sum_{i=1}^{k} x_{Ni}\theta_i}{\partial\theta_k} \end{bmatrix} \tag{5}$$

$$= \begin{bmatrix} x_{11} & x_{21} \dots x_{N1} \\ & \dots \\ x_{1k} & x_{2k} \dots x_{Nk} \end{bmatrix} \tag{6}$$

$$= X^T \tag{7}$$

练习

以（分子）作为主序

- $$\frac{\partial\theta^T X}{\partial\theta} = X$$

证明.

$$\frac{\partial\theta^T X}{\partial\theta} = \begin{bmatrix} \frac{\sum_{i=1}^N x_{i1}\theta_i}{\partial\theta_1} & \frac{\sum_{i=1}^N x_{i2}\theta_i}{\partial\theta_1} \dots \frac{\sum_{i=1}^N x_{ik}\theta_i}{\partial\theta_1} \\ & \dots \\ \frac{\sum_{i=1}^N x_{i1}\theta_i}{\partial\theta_k} & \frac{\sum_{i=1}^N x_{i2}\theta_i}{\partial\theta_k} \dots \frac{\sum_{i=1}^N x_{iN}\theta_i}{\partial\theta_k} \end{bmatrix} \tag{8}$$

$$= \begin{bmatrix} x_{11} & x_{12} \dots x_{1k} \\ & \dots \\ x_{N1} & x_{N2} \dots x_{Nk} \end{bmatrix} \tag{9}$$

$$= X \tag{10}$$

□

练习 I

以（分子）作为主序

证明：

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

，其中 $A_{n\times n} = (a_{ij})_{i=1,j=1}^{n,n}$ 为常数矩阵,x 为 n 维列向量。

练习 II

证明.

$$\frac{\partial x^T A x}{\partial x} = \frac{\begin{aligned}\partial(&a_{11}x_1x_1 + a_{12}x_1x_2 + ... + a_{1n}x_1x_n)\\ &+a_{21}x_2x_1 + a_{22}x_2x_2 + ... + a_{2n}x_2x_n)\\ &+...a_{n1}x_nx_1 + a_{n2}x_nx_2 + ... + a_{nn}x_nx_n)\end{aligned}}{\partial x} = \frac{\partial B}{\partial x}$$

运用元素对列向量求导:

$$= \begin{bmatrix} (a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n) + (a_{11}x_1 + a_{21}x_2 + ... + a_{n1}x_n) \\ (a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n) + (a_{12}x_1 + a_{22}x_2 + ... + a_{n2}x_n) \\ ... \\ (a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n) + (a_{1n}x_1 + a_{2n}x_2 + ... + a_{nn}x_n) \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}\, a_{12} ... a_{1n} \\ a_{21}\, a_{22} ... a_{2n} \\ ... \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \end{bmatrix} + \begin{bmatrix} a_{11}\, a_{21} ... a_{n1} \\ a_{12}\, a_{22} ... a_{n2} \\ ... \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \end{bmatrix} = Ax + A^T x$$

练习

$$f(\beta) = (y - X\beta)^T(y - X\beta)$$

对上式关于 β 求导，其中 y 是 N 维列向量，X 是 $N \times p$ 的矩阵，每行是一个输入向量，假设 X^TX 是非奇异的，求 β 的极值点

练习 I

$$f(\beta) = (y - X\beta)^T(y - X\beta)$$

对上式关于 β 求导，其中 y 是 N 维列向量，X 是 $N \times p$ 的矩阵，每行是一个输入向量，假设 X^TX 是非奇异的，求 β 的极值点

$$\begin{aligned}
(y - X\beta)^T(y - X\beta) &= [y^T - (X\beta)^T][y - X\beta] && (11)\\
&= [y^Ty - (X\beta)^Ty - y^TX\beta + (X\beta)^T(X\beta)] && (12)\\
&= y^Ty - 2y^TX\beta + (X\beta)^TX\beta && (13)\\
& && (14)
\end{aligned}$$

练习 II

$\because \frac{f(\beta)}{\beta} = 0$

$$\begin{aligned} \frac{f(\beta)}{\beta} &= \frac{\partial y^T y - 2y^T X\beta + (X\beta)^T X\beta}{\partial \beta} && (15) \\ &= (2X^T X\beta) - 2X^T y && (16) \\ &= 0 && (17) \end{aligned}$$

令 $A = X^T X$, 则有： $\beta^T A\beta - 2X^T y = 0$

$$2(X^T X)\beta - 2X^T y = 0$$

$$\beta = (X^T X)^{-1} X^T y$$

矩阵的逆矩阵

定义

矩阵的逆矩阵: 如果 A 是一个 $m \times m$ 矩阵, 并且 A 是可逆的, 那么

$$AA^{-1} = A^{-1}A = I$$

- 不可逆的矩阵，我们称为"奇异矩阵"。

课程内容

- 第 2 章迭代优化方法
- 第 3 章监督学习-回归分析
- 第 4 章监督学习-分类算法
- 第 5 章人工神经网络
- 第 6 章非监督学习-聚类与降维分析
- 第 7 章概率图模型-Probabilistic Graphical Model
- 第 8 章遗传算法