chapter 1: 线性回归

符号介绍

输入数据 $X = [x^{(1)}, x^{(2)}, \dots, x^{(n)}]^T$
输出标签 $Y = [y^{(1)}, y^{(2)}, \dots, y^{(n)}]^T$
第i个输入数据 $x^{(i)}$
第i个输出数据 $y^{(i)}$
第i个数据的第j个特征 $x_j^{(i)}(1 <= j <= m)$

概念

- 线性判别分析

知识点

Q1. X必须是满秩的, 倘若不是, 怎么办
A1. 倘若不是满秩的, 则X^TX不存在逆, 无法求解,只能通过正则化限制θ的范围来获得可行解

Q2. 正则化的几何意义以及其公式
A2. 见下文推导过程

Q3. 非参数学习算法与参数学习算法的定义, 区别
A3. 定义: 训练完成后是否需要保存训练数据 区别:非参数学习算法训练完成后不需要保存参数(LR), 而参数学习算法需要(LWR)

Q4. 简述LWR算法过程
A4. 代价函数$J(\theta) = \frac{1}{2}\sum_{i=1}^{i=n} w_i(h_\theta(x^{(i)}) - y^{(i)})^2$其中$w_i = exp(-\frac{(x^{(i)}-x)}{2\gamma^2})$
每当有新数据加入train data时候原数据仍然需要保存

Q5. 二分类问题下逻辑回归的代价函数
A5.

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

$$P(y|x;\theta) = (h_\theta(x^{(i)}))^{y^{(i)}}(1-h_\theta(x^{(i)}))^{1-y^{(i)}}$$

$$J(\theta) = logL(\theta) = \prod_{i=1}^{i=n} logP(y|x,\theta) = \sum_{i=1}^{i=n} y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))$$

> $P(y|x,\theta)$与$P(y|x;\theta)$区别: 前者θ是变量而后者是常量或者给定值

Q6. 牛顿法的求解过程
A6. 牛顿法采用不断迭代$x_{(i+1)} = x_i - \frac{f^{i阶导数}(x_i)}{f^{i+1阶导数}(x_{(i+1)})}$求解出代价函数导数的零点获得最优解

Q7. 从二分类到多分类的几种方法以及其各自的优缺点
A7. 1对其余: 分出单独某一类以及其他类; 1对1: 区分待判定样本是第i类还是第j类(需要$n(n-1)/2$个分类器然后再统计最高票数)

Q8. 简述LDA的算法过程
A8. 面试不常考,在此仅仅给出引用, 西瓜书P60

公式推导以及证明

> 预备知识: (需翻墙)
> 矩阵的迹
> 矩阵求导

- 线性回归问题描述以及采用LMS作为cost function其最优解

  假设函数

  $$h_\theta(x) = \sum_{i=1}^{i=m} \theta_i x_i = \theta^T x$$

  代价函数

  $$J(\theta) = \frac{1}{2}\sum_{i=1}^{i=n}(h_\theta(x^{(i)}) - y^{(i)})^2 = \frac{1}{2}\sum_{i=1}^{i=n}(\theta^T x^{(i)} - y^{(i)})^2$$

  $$\nabla_\theta J(\theta) = \frac{1}{2}\nabla_\theta (X\theta - Y)^T(X\theta - Y)$$

  $$= \frac{1}{2}\nabla_\theta tr(\theta^T X^T X\theta - Y^T X\theta - \theta^T X^T Y + Y^T Y)$$

  $$= \frac{1}{2}\nabla_\theta tr(\theta^T X^T X\theta - 2Y^T X\theta)$$

  $$= \frac{1}{2}\nabla_\theta tr(\theta^T X^T X\theta - 2Y^T X\theta)$$

  $$= \frac{1}{2}tr(X^T X\theta + \theta^T X^T X^T - 2Y^T X)$$

  $$= X^T X\theta + Y^T X$$

  $$= X^T X\theta + X^T Y$$

  令 $\nabla_\theta J(\theta) = 0$可得

  $$\theta = (X^T X)^{-1} X^T Y$$

- 加入L1, L2正则其表达式以及其最优解

  $$J(\theta) = \frac{1}{2}(X\theta - Y)^T(X\theta - Y) + \frac{\lambda}{2}\sum_{j=1}^{j=n}|\theta_j|^q$$

  采用拉格朗日乘子可得

  $$\theta = (X^T X)^{-1} X^T Y \quad \mathrm{s.t.} \sum_{j=1}^{j=n}|\theta_j|^q < \lambda$$

- 使用概率的方法推导LR求解过程 $y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$ 其中$\epsilon^{(i)}$表示经验误差或者非模型影响因素.
  由于$y^{(i)}$的分布与ϵ保持一致,则有

  $$P(y^{(i)}|x^{(i)};\theta) = \frac{1}{\sqrt{2\pi}\sigma}exp(\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2})$$

  采用MLE作为代价函数可得:

$$J(\theta) = L(\theta) = log \prod_{i=1}^{i=n} P(y^{(i)}|x^{(i)};\theta)$$

$$= \sum_{i=1}^{i=n} log \frac{1}{\sqrt{2\pi}\sigma} exp(\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2})$$

$$= nlog \frac{1}{\sqrt{2\pi}\sigma} + \frac{1}{2\sigma^2} \sum_{i=1}^{i=n} (y^{(i)} - \theta^T x)^2$$

> 由上可知, 采用MLE的方法计算LR与LMS做为代价函数是等价的