

习题二

学号, 作者姓名, 邮箱

2017 年 4 月 2 日

1 [10pts] Lagrange Multiplier Methods

请通过拉格朗日乘子法(可参见教材附录B.1)证明《机器学习》教材中式(3.36)与式(3.37)等价。即下面公式(1.1)与(1.2)等价。

$$\begin{aligned} \min_{\mathbf{w}} \quad & -\mathbf{w}^T \mathbf{S}_b \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^T \mathbf{S}_w \mathbf{w} = 1 \end{aligned} \tag{1.1}$$

$$\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w} \tag{1.2}$$

Proof. 此处用于写证明(中英文均可)

□

2 [20pts] Multi-Class Logistic Regression

教材的章节3.3介绍了对数几率回归解决二分类问题的具体做法。假定现在的任务不再是二分类问题，而是多分类问题，其中 $y \in \{1, 2, \dots, K\}$ 。请将对数几率回归算法拓展到该多分类问题。

(1) [10pts] 给出该对率回归模型的“对数似然”(log-likelihood);

(2) [10pts] 计算出该“对数似然”的梯度。

提示1: 假设该多分类问题满足如下 $K - 1$ 个对数几率,

$$\begin{aligned} \ln \frac{p(y=1|\mathbf{x})}{p(y=K|\mathbf{x})} &= \mathbf{w}_1^T \mathbf{x} + b_1 \\ \ln \frac{p(y=2|\mathbf{x})}{p(y=K|\mathbf{x})} &= \mathbf{w}_2^T \mathbf{x} + b_2 \\ &\vdots \\ \ln \frac{p(y=K-1|\mathbf{x})}{p(y=K|\mathbf{x})} &= \mathbf{w}_{K-1}^T \mathbf{x} + b_{K-1} \end{aligned}$$

提示2: 定义指示函数 $\mathbb{I}(\cdot)$,

$$\mathbb{I}(y = j) = \begin{cases} 1 & \text{若 } y \text{ 等于 } j \\ 0 & \text{若 } y \text{ 不等于 } j \end{cases}$$

Solution. 此处用于写解答(中英文均可)

3 [35pts] Logistic Regression in Practice

对数几率回归(Logistic Regression, 简称LR)是实际应用中非常常用的分类学习算法。

(1) [30pts] 请编程实现二分类的LR, 要求采用牛顿法进行优化求解, 其更新公式可参考《机器学习》教材公式(3.29)。详细编程题指南请参见链接: http://lamda.nju.edu.cn/ml2017/PS2/ML2_programming.html

(2) [5pts] 请简要谈谈你对本次编程实践的感想(如过程中遇到哪些障碍以及如何解决, 对编程实践作业的建议与意见等)。

Solution. 此处用于写解答(中英文均可)

4 [35pts] Linear Regression with Regularization Term

给定数据集 $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, 其中 $\mathbf{x}_i = (x_{i1}; x_{i2}; \dots; x_{id}) \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, 当我们采用线性回归模型求解时, 实际上是在求解下述优化问题:

$$\hat{\mathbf{w}}_{\text{LS}}^* = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2, \quad (4.1)$$

其中, $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^m$, $\mathbf{X} = [\mathbf{x}_1^T; \mathbf{x}_2^T; \dots; \mathbf{x}_m^T] \in \mathbb{R}^{m \times d}$, 下面的问题中, 为简化求解过程, 我们暂不考虑线性回归中的截距(intercept)。

在实际问题中, 我们常常不会直接利用线性回归对数据进行拟合, 这是因为当样本特征很多, 而样本数相对较少时, 直接线性回归很容易陷入过拟合。为缓解过拟合问题, 常对公式(4.1)引入正则化项, 通常形式如下:

$$\hat{\mathbf{w}}_{\text{reg}}^* = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \Omega(\mathbf{w}), \quad (4.2)$$

其中, $\lambda > 0$ 为正则化参数, $\Omega(\mathbf{w})$ 是正则化项, 根据模型偏好选择不同的 Ω 。

下面, 假设样本特征矩阵 \mathbf{X} 满足列正交性质, 即 $\mathbf{X}^T \mathbf{X} = \mathbf{I}$, 其中 $\mathbf{I} \in \mathbb{R}^{d \times d}$ 是单位矩阵, 请回答下面的问题(需要给出详细的求解过程):

- (1) [5pts] 考虑线性回归问题, 即对应于公式(4.1), 请给出最优解 $\hat{\mathbf{w}}_{\text{LS}}^*$ 的闭式解表达式;
- (2) [10pts] 考虑岭回归(ridge regression)问题, 即对应于公式(4.2)中 $\Omega(\mathbf{w}) = \|\mathbf{w}\|_2^2 = \sum_{i=1}^d w_i^2$ 时, 请给出最优解 $\hat{\mathbf{w}}_{\text{Ridge}}^*$ 的闭式解表达式;
- (3) [10pts] 考虑LASSO问题, 即对应于公式(4.2)中 $\Omega(\mathbf{w}) = \|\mathbf{w}\|_1 = \sum_{i=1}^d |w_i|$ 时, 请给出最优解 $\hat{\mathbf{w}}_{\text{LASSO}}^*$ 的闭式解表达式;
- (4) [10pts] 考虑 ℓ_0 -范数正则化问题,

$$\hat{\mathbf{w}}_{\ell_0}^* = \arg \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_0, \quad (4.3)$$

其中, $\|\mathbf{w}\|_0 = \sum_{i=1}^d \mathbb{I}[w_i \neq 0]$, 即 $\|\mathbf{w}\|_0$ 表示 \mathbf{w} 中非零项的个数。通常来说, 上述问题是NP-Hard问题, 且是非凸问题, 很难进行有效地优化得到最优解。实际上, 问题(3)中的LASSO可以视为是近些年研究者求解 ℓ_0 -范数正则化的凸松弛问题。

但当假设样本特征矩阵 \mathbf{X} 满足列正交性质, 即 $\mathbf{X}^T \mathbf{X} = \mathbf{I}$ 时, ℓ_0 -范数正则化问题存在闭式解。请给出最优解 $\hat{\mathbf{w}}_{\ell_0}^*$ 的闭式解表达式, 并简要说明若去除列正交性质假设后, 为什么问题会变得非常困难?

Solution. 此处用于写解答(中英文均可)