软件工程本科生《机器学习》课程教案（七）

讲解人： 李济洪 (教授)、王瑞波 (讲师)

一、授课课题

线性模型选择及正则化

二、授课时间

2019 年 10 月 22 日星期二 8:00am-10:00am

三、课时安排

2 课时

四、授课类型

理论课

五、教材

加雷斯 · 詹姆斯, 丹妮拉 · 威滕, 等.《统计学习导论: 基于 R 应用》[M]. 机械工业出版社, 2015.

课程网站：`http://www-bcf.usc.edu/~gareth/ISL/`

作业网站：`https://github.com/nguyen-toan/ISLR`

六、对应章节

第六章

七、教学目标及要求

1. 掌握算法风险估计的多种方法，包括：C_p, AIC, BIC 及调整 R^2

八、教学重点

1. 掌握子集选择方法；
2. 掌握岭回归及 Lasso 回归方法；

九、教学难点

1. 岭回归及 Lasso 回归的基本思想；
2. 风险估计各种方法的区分；

十、教学方式

讲授

十一、教学手段

课件讲解 + 课间讨论

十二、教学过程

1. 上讲回顾（教学方式：讲授；时间：5 分钟）

上讲内容主要讲授了贝叶斯概念学习的原理，使用数字游戏（Number game）揭示了基于概率机制的概念学习过程，其中，具体讲解了假设（hypotheses）、似然、先验分布、后验分布、极大似然估计、极大后验估计、后验预测分布等概念。最后，简要介绍了二项分布的共轭分布为 Beta 分布，以及多项分布的共轭分布为狄利克雷分布。

2. 引入新课（教学方式：讲授；时间：5 分钟）

介绍本讲的主要关注点在于：给定 p 个预测变量，如何选取出一个变量集合，使得所生成的多元线性回归具有高的预测性能及好的解释性能。

3. 介绍算法预测性能的估计方法（教学方式：讲授；时间：20 分钟）

1. 介绍预测性能估计的基本问题；
2. 介绍模型复杂度、方差-偏差权衡以及过拟合等概念；
3. 介绍交叉验证估计的缺点；
4. 介绍使用经验风险估计来构造预测性能估计的几种方法。
   - 介绍 C_p 估计；
   - 介绍 AIC 准则；
   - 介绍 BIC 准则；

4. 介绍子集选择方法（教学方式：讲授；时间：20 分钟）

4.1. 介绍子集选择的基本问题

对于多元线性回归算法，子集选择的基本问题是：从 p 个候选预测变量中，选择重要的变量，提升算法预测性能。该基本问题对应于如下优化问题：

$$\begin{aligned} &\min_{\beta}\{\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2\} \\ s.t.\quad &\sum_{j=1}^{p} I(\beta_j \neq 0) \leq s \end{aligned} \tag{1}$$

4.2 介绍最优子集选择

最优子集选择的**基本思想**：对 p 的预测变量的所有可能组合分别使用最小二乘回归进行拟合，然后，依据预测性能的估计，从所有可能模型中选择一个最优模型。

给出最优子集选择算法，并介绍最优子集选择的优点和缺点。

4.2 介绍逐步选择

- 介绍前向选择的基本思想和优缺点。
- 介绍后项选择的基本思想和优缺点。
- 介绍逐步选择方法；

5 介绍压缩估计方法（教学方法：讲授；时间：30 分钟）

5.1 介绍原始优化问题的另一种解法

分析原始优化问题的解决难点。对约束条件进行放宽，得到原始问题的近似问题。为此，需要思考如下问题：

如何对原始问题的约束条件进行放宽？

结合凸优化问题，可以将原始问题中的 l_0 约束条件放宽至 l_1 范数和 l_2 范数。

5.2 介绍岭回归

1. 介绍岭回归对应的带约束优化问题。然后，使用拉格朗日乘子法转换该问题。进而，给出岭回归估计的具体形式。
2. 对比最小二乘估计和岭估计的具体形式，进而讲解岭估计的特点：
    - 有偏估计
    - 方差小、较稳定

3. 介绍岭估计的 MSE 与超参数 λ 之间的关系。
    - λ 变大，偏差增大，方差减小，MSE 先减后增。
4. 给出岭回归在 Credit 数据集上的应用。

5.3 介绍 Lasso 回归

1. 介绍 Lasso 回归对应的约束优化问题，及拉格朗日乘子法对应的等价优化问题。
2. 介绍 lasso 估计的特点。
    - 向 0 的方向压缩；
    - 当 λ 足够大时，Lasso 估计精确为 0。
    - Lasso 估计具有稀疏性。
    - Lasso 估计可用于解释变量的重要性。
3. 介绍 Lasso 回归在 Credit 数据集上的应用。

5.4 对比最优子集回归、岭回归和 Lasso 回归

1. 从优化问题的约束条件角度，解释最优子集回归、岭回归和 Lasso 回归之间的关系。
2. 基于二变量的情形，对比岭回归和 Lasso 回归的约束区域，并解释 Lasso 估计解稀疏性的原因。
3. 以 45 个变量的模拟数据情形，分别考虑真变量为 45 和 2 两种情形下，岭回归和 Lasso 回归的优劣。
4. 在下述特殊情形下，对比岭估计和 Lasso 估计的形式异同。
    - $n = p$ 且 $X = I$，此时，残差平方和为 $RSS = \sum_{j=1}^{p}(y_j - \beta_j)^2$。

5.5 给出岭回归和 Lasso 回归的贝叶斯后验解释

从后验估计的角度，给出岭回归、Lasso 回归与最小二乘回归之间的关系。

$$p(\beta|X,Y) \propto f(Y|X,\beta)p(\beta|X) = f(Y|X,\beta)p(\beta) \tag{2}$$

其中，密度 $f(Y|X,\beta)$ 对应于最小二乘线性回归。岭回归和 Lasso 回归的差别在于 $p(\beta)$ 的选取不同。其中，

- 岭回归: $p(\beta)$ 为高斯分布。
- Lasso 回归：$p(\beta)$ 为拉普拉斯分布。

5.6 介绍超参数选择方法

岭回归和 Lasso 回归中含有一个超参数 λ。该参数用于调节估计的偏度和方差，进而产生不同预测性能的估计值。λ 的选取通常采用交叉验证方法。本节中，仅给出：

- 基于留一交叉验证估计的岭回归超参数 λ 的选取实验。
- 基于 10 折交叉验证的 Lasso 回归超参数 λ 的选取实验。

但，如下问题需要进一步思考：

1. 何种交叉验证方法可以得到更为可靠的 λ 的估计值?
2. 当模型中含有多组超参数时，如何使用交叉验证方法进行选择?

十三、作业

1.【推导】岭回归的参数估计形式。
2.【思考】岭回归估计与极大后验估计的关系。
3.【编程】岭回归方法和 Lasso 回归方法在 Credit 数据集上的应用实现。

十四、参考资料