机器学习导论
习题六

151242041, 王昊庭, hatsuyukiw@gmail.com

2017 年 6 月 9 日

1 [20pts] Ensemble Methods

(1) **[10pts]** 试说明 Boosting 的核心思想是什么，Boosting 中什么操作使得基分类器具备多样性？

(2) **[10pts]** 试析随机森林为何比决策树 Bagging 集成的训练速度更快。

Solution. (1) Boosting 的核心思想是先从初始训练集训练出一个基学习器, 之后根据基学习器的表现, 改变训练集, 使得后续的学习器可以将重点放在之前的学习器的偏差上. 改变训练集的主要方法有

- 改变样本分布, 使得之前的学习器分类错误或者误差较大的的样本的概率增大, 使得后续分类器更加关注这类样本, 比如 AdaBoost 使用的就是这个方法
- 后续训练器在之前的学习器的的误差上进行学习, 常见于回归模型

等.

(2) 随机森林引入了属性随机选择的特性, 单个决策树考虑的属性减少, 速度自然比考虑全部属性的决策树更快. 故而在决策树数目相同的情况下, 随机森林比决策树 Bagging 集成的训练速度更快.

2 [20pts] Bagging

考虑一个回归学习任务 $f:\mathbb{R}^d \to \mathbb{R}$。假设我们已经学得 M 个学习器 $\hat{f}_1(\mathbf{x}), \hat{f}_2(\mathbf{x}), \ldots, \hat{f}_M(\mathbf{x})$。我们可以将学习器的预测值看作真实值项加上误差项

$$\hat{f}_m(\mathbf{x}) = f(\mathbf{x}) + \epsilon_m(\mathbf{x}) \tag{2.1}$$

每个学习器的期望平方误差为 $\mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})^2]$。所有的学习器的期望平方误差的平均值为

$$E_{av} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})^2] \tag{2.2}$$

M 个学习器得到的 Bagging 模型为

$$\hat{f}_{bag}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M}\hat{f}_m(\mathbf{x}) \tag{2.3}$$

Bagging 模型的误差为

$$\epsilon_{bag}(\mathbf{x}) = \hat{f}_{bag}(\mathbf{x}) - f(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M}\epsilon_m(\mathbf{x}) \tag{2.4}$$

其期望平均误差为

$$E_{bag} = \mathbb{E}_{\mathbf{x}}[\epsilon_{bag}(\mathbf{x})^2] \tag{2.5}$$

(1) **[10pts]** 假设 $\forall\, m \neq l$, $\mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})] = 0$, $\mathbb{E}_{\mathbf{x}}[\epsilon_m(\mathbf{x})\epsilon_l(\mathbf{x})] = 0$。证明

$$E_{bag} = \frac{1}{M}E_{av} \tag{2.6}$$

(2) **[10pts]** 试证明不需对 $\epsilon_m(\mathbf{x})$ 做任何假设, $E_{bag} \leq E_{av}$ 始终成立。(提示：使用 Jensen's inequality)

Proof. (1)

$$\begin{aligned}
E_{bag} &= \mathbb{E}_{\mathbf{x}}[\epsilon_{bag}(\mathbf{x})^2] \\
&= \frac{1}{M^2}\mathbb{E}_{\mathbf{x}}[(\sum_{m=1}^{M}\epsilon_m(\mathbf{x}))^2] \\
&= \frac{1}{M^2}\left(\sum_{i=1}^{M}\mathbb{E}_{\mathbf{x}}[(\epsilon_i(\mathbf{x}))^2] + \sum_{i=1,j=2,i\neq j}^{M}\mathbb{E}_{\mathbf{x}}[\epsilon_i(\mathbf{x})\epsilon_j(\mathbf{x})]\right) \\
&= \frac{1}{M}\left(\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}_{\mathbf{x}}[(\epsilon_i(\mathbf{x}))^2]\right) \\
&= \frac{1}{M}E_{av}\,.
\end{aligned}$$

(2) The function $f(x) = x^2$ is convex. Thus Jensen's inequality $f(\frac{1}{N}\sum_{i=1}^{N}x_i) \leq \sum_{i=1}^{N}\frac{1}{N}f(x_i)$ holds.

$$\begin{aligned}
E_{bag} &= \mathbb{E}_{\mathbf{x}}[\epsilon_{bag}(\mathbf{x})^2] \\
&= \frac{1}{M}\mathbb{E}_{\mathbf{x}}[M(\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(\mathbf{x}))^2] \\
&\leq \frac{1}{M}\mathbb{E}_{\mathbf{x}}[\sum_{m=1}^{M}(\epsilon_m(\mathbf{x}))^2] \\
&= \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}_{\mathbf{x}}[(\epsilon_m(\mathbf{x}))^2]) \\
&= E_{av}\,.
\end{aligned}$$

□

3 [30pts] AdaBoost in Practice

(1) **[25pts]** 请实现以 Logistic Regression 为基分类器的 AdaBoost,观察不同数量的 ensemble 带来的影响。详细编程题指南请参见链接：http://lamda.nju.edu.cn/ml2017/PS6/ML6_programming.html

(2) **[5pts]** 在完成上述实践任务之后，你对 AdaBoost 算法有什么新的认识吗？请简要谈谈。

Solution. 这个作者认为本题的算法在理论上是无意义的, 在实践中是错误的.

LR 的本质是给出一个空间中的超平面划分, AdaBoost 将若干个分类器线性组合, 故而 LR AdaBoost 的最终决策界仍然是空间中的超平面. 也就是说 LR AdaBoost 的假设空间位于单个 LR 的假设空间内, 故而单个 LR 最优假设优于或等于 LR AdaBoost 最优假设.

在无正则项的情形下, 由于 LR 是凸优化问题, 很容易得到 LR 的假设空间中的这个最优假设. 另一方面, AdaBoost 或许可以视为一种正则, 但是从实验来看 LR 的 L2 regularization 完全优于 AdaBoost 的正则 (这个作者认为在 LR 上 AdaBoost 实际没有正则的效果).

总之, 无论是从理论还是从实践的角度分析, LR AdaBoost 都完全没有意义.