HW8

PB20111689 蓝俊玮

习题 8.2

对于任意损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 来说有
$E_{\boldsymbol{x}}(l(-f(\boldsymbol{x})H(\boldsymbol{x}))) = l(-H(\boldsymbol{x}))P(f(\boldsymbol{x})=1|\boldsymbol{x}) + l(H(\boldsymbol{x}))P(f(\boldsymbol{x})=-1|\boldsymbol{x})$，而当 $P(f(\boldsymbol{x})=1|\boldsymbol{x}) > P(f(\boldsymbol{x})=-1|\boldsymbol{x})$ 时，为了能够让损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 更小，我们希望得到 $l(-H(\boldsymbol{x})) < l(H(\boldsymbol{x})) = l(-(-H(\boldsymbol{x})))$，因为损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 对 $H(\boldsymbol{x})$ 来说是递减的，因此希望得到 $H(\boldsymbol{x}) > -H(\boldsymbol{x})$ 即 $H(\boldsymbol{x}) = 1$，这样就可以满足 $l(-1) < l(-(-1))$；同理当 $P(f(\boldsymbol{x})=1|\boldsymbol{x}) < P(f(\boldsymbol{x})=-1|\boldsymbol{x})$ 时，为了能够让损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 更小，我们希望得到 $l(-H(\boldsymbol{x})) > l(H(\boldsymbol{x})) = l(-(-H(\boldsymbol{x})))$，即希望得到 $H(\boldsymbol{x}) < -H(\boldsymbol{x})$ 即 $H(\boldsymbol{x}) = -1$，这样可以满足 $l(-(-1)) > l(-1)$。

因此从上述我们可以得知：

$$H(\boldsymbol{x}) = \begin{cases} 1, & P(f(\boldsymbol{x})=1|\boldsymbol{x}) > P(f(\boldsymbol{x})=-1|\boldsymbol{x}) \\ -1, & P(f(\boldsymbol{x})=1|\boldsymbol{x}) < P(f(\boldsymbol{x})=-1|\boldsymbol{x}) \end{cases}$$

所以意味着 $H(\boldsymbol{x})$ 达到了贝叶斯最优错误率，换言之，若损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 最小化，意味着分类错误率也最小化。说明该损失函数 $l(-f(\boldsymbol{x})H(\boldsymbol{x}))$ 是 0-1 损失函数的一致替代函数。

习题 8.8

MultiBosoting 能够利用 AdaBoost 大量降低偏差和方差以及 wagging 显著地减少方差来有效地降低偏差和方差。 同时通过使用 C4.5 作为基础学习算法，MultiBoosting 可以产生比 AdaBoost 错误率更低的决策。但是其训练成本和预测成本都会显著增加，且MultiBosoting 可能会加重过拟合问题。
Iterative Bagging 可以降低方差，且可以解决过拟合问题。由于 Bagging 本身就是一种降低方差的算法，所以 Iterative Bagging 相当于Bagging与单分类器的折中，且它无法降低偏差。