集成学习

集成学习和方差大的模型组合效果较好

输入x和输出y真实 $y=h(x)$

对于M个不同模型 $f_i(x)$

$$R(f_m)=E_x[(f_m(x)-h(x))^2]$$
$$=E_x[\epsilon_m(x)^2]$$

$\epsilon_m(x)=f_m(x)-h(x)$ 是模型m在样本x上的错误

$$\bar{R}(f)=\frac{1}{M}\sum_{m=1}^{M}E_x[\epsilon_m(x)^2]$$

通过某种策略集成多个模型，降低决策来提高准确率

- 直接平均
- 加权平均

投票：$F(x)=\frac{1}{M}\sum_{m=1}^{M}f_m(x)$

$$R(F)=E_x[(\frac{1}{M}\sum_{m=1}^{M}f_m(x)-h(x))^2]$$
$$=E_x[(\frac{1}{M}\sum_{m=1}^{M}\epsilon_m(x))^2]$$
$$=\frac{1}{M^2}E_x[\sum_{m=1}^{M}\sum_{n=1}^{M}\epsilon_m(x)\epsilon_n(x)]$$
$$=\frac{1}{M^2}\sum_{m=1}^{M}\sum_{n=1}^{M}E_x[\epsilon_m(x)\epsilon_n(x)]$$

其中 $E_x[\epsilon_m(x)\epsilon_n(x)]$ 是两个模型错误的相关性

不相关 $E_x[\epsilon_m(x)\epsilon_n(x)]=0\ \forall m\neq n$

相同 $\epsilon_m(x)=\epsilon_n(x)\ \forall m\neq n$

且 $\epsilon_m\geq 0\ \forall m$ 得

$$\bar{R}(f)\geq R(F)\geq\frac{1}{M}\bar{R}(f)$$

要求每个模型有差异性，随着模型数量增多，错误率下降趋于0

Bagging类

随机初始训练样本，随机选择特征提高每个模型的独立性

①Bagging (Bootstrap Aggregating)：

原始训练集上进行有放回的采样(随机)得到M个比较小的训练集训练M个模型

然后投票

②随机森林

引入随机特征

基模型是决策树

Boosting类

先后训练不同的基模型，每个模型针对前序模型的错误进行专门训练。

根据前序模型的结果，来调整训练样本的权重。从而增加差异性

①AdaBoost.

加法模型

$$F(x)=\sum_{m=1}^{M}\underbrace{\alpha_m}_{\text{强分类器}}\underbrace{f_m(x)}_{\text{弱分类器}}$$

学习到第m个弱分类器后，增加该分样本的权重，使得第m+1个弱分类器更关注于分错的样本

分步迭代加性模型

$L(F) = \exp(-yF(x))$

$= \exp(-y\sum_{m=1}^{M}\alpha_m f_m(x))$

$y, f_m(x) \in \{+1, -1\}$

假设经过$m-1$次迭代

$F_{m-1}(x) = \sum_{t=1}^{m-1}\alpha_t f_t(x)$

第m次迭代目标找一个$\alpha_m, f_m(x)$使得

$L(\alpha_m, f_m(x)) = \sum_{n=1}^{N}\exp(-y^n(F_{m-1}(x^n) + \alpha f_m(x^n)))$ 最小

令$w_m^{(n)} = \exp(-y^n F_{m-1}(x^n))$

则

$L(\alpha_m, f_m(x)) = \sum_{n=1}^{N} w_m^{(n)}\exp(-y^n\alpha_m f_m(x^n))$

$\because y, f_m(x) \in \{+1, -1\}$,

$\therefore yf_m(x) = 1 - 2I(y \neq f_m(x))$

将损失函数在$-\alpha_m y^n f_m(x^n)$处展开

$L(\alpha_m, f_m(x)) = \sum_{n=1}^{N} w_m^{(n)}(1 - \alpha_m y^{(n)} f_m(x^{(n)}) + \frac{1}{2}\alpha_m^2)$

$\propto \alpha_m \sum_{n=1}^{N} w_m^{(n)} I(y^{(n)} \neq f_m(x^{(n)}))$

当$\alpha_m > 0$时，最小分类器是使样本权重为$w_m^{(n)}$时加权错误率最小分类器

$L(\alpha_m, f_m(x)) = \sum_{y^{(n)}=f_m(x)} w_m^{(n)}\exp(-\alpha_m) + \sum w_m^{(n)}\exp(\alpha_m)$

$\propto (1-\epsilon_m)\exp(-\alpha_m) + \epsilon_m\exp(\alpha_m)$

$f_m(x)$加权错误率

$\epsilon_m = \frac{\sum_{y^{(n)} \neq f_m(x_m^n)} w_m^{(n)}}{\sum_n w_m^{(n)}}$

$\alpha_m = \frac{1}{2}\log\frac{1-\epsilon_m}{\epsilon_m}$