Bagging & Random forest

Bagging: Boostrap AGGregretion: ①: Sample bragging

Key: Repeatedly sample (with replacement!) a collection of training examples and train a model on that sample

流程: for $t = 1, \cdots, T$ do

for $s = 1, \cdots, S$ do 注意，是对每一个 h_t 都重采样一份 S_t

$i_s \sim \text{Uniform}(1, \cdots, N)$

$S_t = \{(x^{(i_s)}, y^{(i_s)})\}_{s=1}^{S}$ ← boostrap sample.

$h_t = \text{train}(S_t)$. ← classifier

return $\hat{h}(x) = \text{aggregate}^{*}(h_1, \cdots, h_T)$

* for classification: majority vote

for regression: average

② Feature Bagging key: select subset of feature as well

for $t = 1, \cdots, T$ do

for $s = 1, \cdots, S$ do

$m_s \sim \text{Uniform}(1, \cdots, M)$

for $i = 1, \cdots, N$ do

$\tilde{x}^{(i)} = [x_{m_1}^{(i)}, x_{m_2}^{(i)}, \cdots, x_{m_S}^{(i)}]^T$. ← sample feature

$D_t = \{(\tilde{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$ ← subspace

$h_t = \text{train}(D_t)$

return $\hat{h}(x) = \text{aggregate}(h_1, \cdots, h_T)$

Random Forest: Key: Combine prediction of many diverse decision trees to reduce variability

If B r.v. all have variance σ^2, Then $\frac{1}{B}\sum_{b=1}^{B} x^{(b)}$'s variance is: σ^2/B !

No.

Date

So we can combine (sample) bagging and a specific variant of feature bagging to train decision trees

基决策树

流程：重复：抽一个样本（with replacement）

→ 对树的每个结点，先从该结点属性集合中随机选一个子集，再在子集中选最优属性。设|集合| = d，|子集| = k，若 k=d：则 vanilla decision tree splitting node；若 k=1，则随机选一个作结点

建议：$k = \log_2 d$

上述算法中的采样有一个优点：每个基学习器只用了dataset中63.2%的样本，剩下36.8%*可用作 validation set 来对泛化能力作"包外估计"(out of bag estimate)；这便可用于超参优化！

Random forest 简单易实现，且 performance 很不错，而且收敛性与 Bagging 相似

*: $\lim_{m\to\infty}\left(1-\frac{1}{m}\right)^m \approx 0.368$