STAT430: Machine Learning for Financial Data

LarryHua.com/teaching

Spring 2019

Ensemble methods

Bootstrap aggregation (bagging)

- Reduce variances in forecasts
- Steps:
  1. generate N training datasets by random sampling **with** replacement
  2. fit N estimators, one on each training set
  3. the ensemble forecast is the simple average of the individual forecasts from the N models

Bagging - reduce variance

- Bagging's main advantage: reduce forecasts' variance, hence helping address overfitting
  - Let φ_i, $i = 1, \ldots, N$, be the predictions, and define the average variance $\bar{\sigma}^2 = (\sum_{i=1}^{N} \sigma_i^2)/N$, and the **average correlation** $\bar{\rho} := \left(\sum_{j \neq i} \sigma_i \sigma_j \rho_{ij}\right) / \left(\bar{\sigma}^2 N(N-1)\right)$, then

$$
\begin{aligned}
V\left((1/N)\sum_{i=1}^{N}\varphi_i\right) &= (1/N^2)\left(\sum_{i=1}^{N}\sigma_i^2 + \sum_{i \neq j}\sigma_i\sigma_j\rho_{i,j}\right) \\
&= \bar{\sigma}^2\left(\bar{\rho} + \frac{1-\bar{\rho}}{N}\right).
\end{aligned}
$$

  - Reduce $\bar{\rho}$ to reduce the variance of predictions; when N is large, $(1-\bar{\rho})/N$ diminishes.

Bagging - Limitations

- Consider k classes, n independent classifiers, and p the probability of correct prediction.
- Let S_n be the total number of success, then $S_n \sim \mathrm{Binomial}(n, p)$.
- The chance of being better than **random guess** is $P[S_n > n/k] = \sum_{i=\lfloor n/k \rfloor + 1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$
- By WLLN, $p > 1/k$ implies that $\lim_{n\to\infty} P[S_n/n > 1/k] = \lim_{n\to\infty} P[S_n/n - p > 1/k - p] \to 1$. Therefore, when n is large enough, we can have $P[S_n > n/k] > p$.

Bagging - Limitations

- Bagging is more likely to be successful in reducing variance than in reducing bias.
- If the individual learners are poor classifiers (i.e., $p \ll 1/k$, assuming k classes), majority voting will still perform poorly (although with lower variance).
  - It's relative easier to have $\bar{\rho} \ll 1$ than $p > 1/k$
- See Figure 6.2 in AFML

Redundancy in financial data

- Samples drawn **with** replacement are more likely to be virtually identical, even if they do not share the same observations.
  - When $\bar{\rho} \approx 1$, $V((1/N)\sum_{i=1}^{N} \varphi_i)$ can not be reduced.
- Inflate out-of-bag accuracy.

Redundancy in financial data - solutions

- An easy solution: set a maximum percentage of samples to be used each time for bootstrap used in bagging
  - the maximum percentage can be the mean of the average uniqueness of the labels, i.e., $(1/I) \sum_{i=1}^{I} \bar{u}_i$.
  - rationale: not sample more frequently than the uniqueness
- A better solution: sequential bootstrap for bagging
  - No information leakage even if use info from the test set to calculate uniqueness for features bars in the training set
  - However, labels between training and test sets may overlap (more to discuss in the cross validation section)
  - Can be very time consuming, and require parallel computing in R
- Try R

Random forest

- Like bagging, RF reduces forecasts' variance without overfitting
- Improve the variance reduction of bagging by reducing the correlation between the trees, without increasing the variance too much
- RF evaluates feature importance
- Like bagging, RF will not necessarily exhibit lower bias than individual decision trees
- Sequential bootstrap can also be applied for random forest

Boosting

- Typical steps for adaBoost

1. generate one training set by random sampling **with** replacement, according to some sample weights (initialized with uniform weights).
2. fit one estimator using that training set.
3. if the single estimator achieves an accuracy greater than the acceptance threshold (better than guess), the estimator is kept, otherwise it is discarded.
4. give more weight to mis-classified observations, and less weight to correctly classified observations.
5. repeat the previous steps until N estimators are produced.
6. the ensemble forecast is the weighted average of the individual forecasts from the N models, where the weights are determined by the accuracy of the individual estimators.

Boosting in finance

- Differences from bagging
  - Individual classifiers are fit sequentially
  - Poor-performing classifiers are dismissed
  - Observations are weighted differently in each iteration
  - The ensemble forecast is a weighted average of the individual learners
- Pros and cons
  - Reduces both variance and bias in forecasts, but greater risk of overfitting
  - **For financial data, bagging is preferred**
  - Bagging can be paralleled, but boosting requires sequential running (can do parallel in building trees, e.g., xgboost)

Boosting - some R functions to consider

- Adaptive Boosting (Adaboost)
  - R: adabag::boosting
- Extreme Gradient Boosting (Xgboost)
  - R: xgboost::xgboost
- Try R
- Back to Course Scheduler