ML Lecture 8 10/26/11

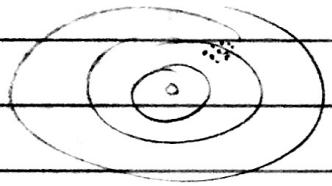
Last time	Today	Next time
• Info Theo. Primer	• Finish Dec. trees	• Online learning
• Decision trees	• Ensemble methods	- Halving algorithm
	- boosting	- Hedge
	- bagging	

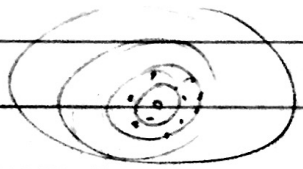
---

bias/variance tradeoff : analogous to accuracy/precision

$MSE = var + bias^2$

bullseye





precise, but not accurate
(high bias, low variance)

accurate, but not precise
(low bias, high variance)

- low complexity models:
  inaccurate but not variable

- high complexity models:
  accurate but variable, depending on train; data

How to get best of both worlds?

① bagging: eliminate variance of accurate models
- average over lots of learners, trained over many random data sets

② boosting:
- average lots of learners trained over carefully chosen data sets
- reduces bias & variance
  ↑ carefully chosen dataset ↑ averaging

Bootstrap aggregating

From Wikipedia, the free encyclopedia

Bootstrap aggregating (**bagging**) is a machine learning ensemble meta-algorithm to improve machine learning of classification and regression models in terms of stability and classification accuracy. It also reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree models, it can be used with any type of model. Bagging is a special case of the model averaging approach.

Contents

- n objects
- n samples w/ replacement
- prob miss item i w/ all n samples

  $(1-1/n)^n \to 1/e \approx 0.368$
- prob hit item i $\to 1-1/e \approx 0.632$
- expected # hits $= n\cdot(1-1/e)$

  $\approx n \cdot 0.632$

63.2% of data

Description of the technique

Given a standard training set D of size n, bagging generates m new training sets D_i, each of size $n' \le n$, by sampling examples from D uniformly and with replacement. By sampling with replacement, it is likely that some examples will be repeated in each D_i. If $n'=n$, then for large n the set D_i is expected to have 63.2% of the unique examples of D, the rest being duplicates. This kind of sample is known as a bootstrap sample. The m models are fitted using the above m bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

Since the method *averages* several predictors, it is not useful for improving *linear* models. Similarly, bagging does not improve very stable models like k nearest neighbors.

Example: Ozone data

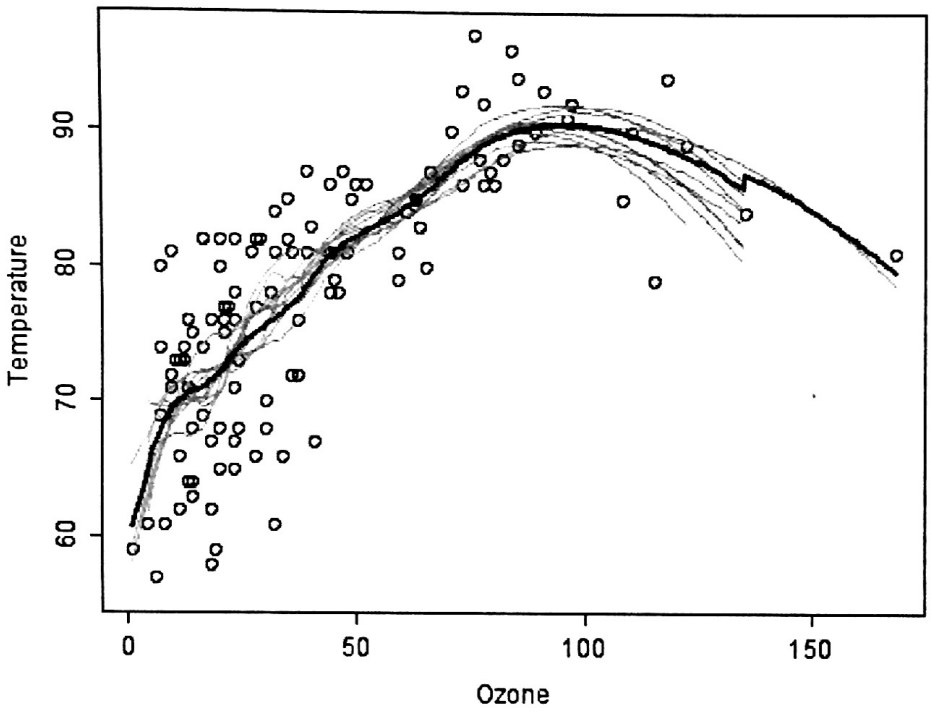
This example is rather artificial, but illustrates the basic principles of bagging.

Rousseeuw and Leroy (1986) describe a data set concerning ozone levels. The data are available via the classic data sets page. All computations were performed in R.

A scatter plot reveals an apparently non-linear relationship between temperature and ozone. One way to model the relationship is to use a loess smoother. Such a smoother requires that a span parameter be chosen. In this example, a span of 0.5 was used.

One hundred bootstrap samples of the data were taken, and the LOESS smoother was fit to each sample. Predictions from these 100 smoothers were then made across the range of the data. The first 10 predicted smooth fits appear as grey lines in the figure below. The lines are clearly very *wiggly* and they overfit the data - a result of the span being too low.

The red line on the plot below represents the mean of the 100 smoothers. Clearly, the mean is more stable and there is less overfit. This is the bagged predictor.



History

Bagging (**B**ootstrap **agg**rega**ting**) was proposed by Leo Breiman in 1994 to improve the classification by combining classifications of randomly generated training sets. See Breiman, 1994. Technical Report No. 421.

See also

- Boosting
- Cross validation

References

- Leo Breiman (1996). "Bagging predictors" (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.7654&rep=rep1&type=pdf) . *Machine Learning* **24** (2): 123–140. doi:10.1007/BF00058655 (http://dx.doi.org/10.1007%2FBF00058655) . http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.7654&rep=rep1&type=pdf.

Categories: Ensemble learning | Machine learning | Computational statistics | Artificial intelligence stubs | Statistics stubs