Data Science

Ensembles Continued

Review Last Time

- What is an ensemble?
- Bootstrap
- Bagging
- Random Forest

Today

- Boosting
- Stacked Generalization

Boosting

- Another ensemble method

Boosting

- Another ensemble method
- On the surface looks similar to bagging/RF

Boosting

- Another ensemble method
- On the surface looks similar to bagging/RF
- Different history: Boosting was created by theorists
  - Robert Shapire and Yoav Freund
  - Other important people: Michael Kearns and Leslie Valiant

Boosting

- Another ensemble method
- On the surface looks similar to bagging/RF
- Different history: Boosting was created by theorists
  - Robert Shapire and Yoav Freund
  - Other important people: Michael Kearns and Leslie Valiant
- Question: Can we make a set of weak learners into a strong learner?

Boosting

- Another ensemble method
- On the surface looks similar to bagging/RF
- Different history: Boosting was created by theorists
  - Robert Shapire and Yoav Freund
  - Other important people: Michael Kearns and Leslie Valiant
- Question: Can we make a set of weak learners into a strong learner?
  - What are weak and strong learners?

Boosting

- Q: What's a weak learner? Thoughts? (Classification)

Boosting

- Q: What's a weak learner? Thoughts? (Classification)
- A: A classifier that does only slightly better than random. It's accuracy is 50% + e where e is a small number.

Boosting

- Q: What's a weak learner? Thoughts? (Classification)
- A: A classifier that does only slightly better than random. It's accuracy is 50% + e where e is a small number.
- Q: What's a strong learner? Thoughts?

Boosting

- Q: What's a weak learner? Thoughts? (Classification)
- A: A classifier that does only slightly better than random. It's accuracy is 50% + e where e is a small number.
- Q: What's a strong learner? Thoughts?
- A: A classifier that does very well. Accuracy 100% - e

Boosting

- Q: What's a weak learner? Thoughts? (Classification)
- A: A classifier that does only slightly better than random. It's accuracy is 50% + e where e is a small number.
- Q: What's a strong learner? Thoughts?
- A: A classifier that does very well. Accuracy 100% - e
- Question: Can we make a single strong learner out of a set of weak learners?

Boosting

- Boosting attempts to combine weak learners into strong learners

Boosting

- Boosting attempts to combine weak learners into strong learners
- How is this different than Bagging?

Boosting

- Boosting attempts to combine weak learners into strong learners
- How is this different than Bagging?
  - Q: What are we doing when we try bagging?

Boosting

- Boosting attempts to combine weak learners into strong learners
- How is this different than Bagging?
  - Q: What are we doing when we try bagging?
  - A: Trying to average out variance. Reduce variance

Boosting

- Boosting attempts to combine weak learners into strong learners
- How is this different than Bagging?
  - Q: What are we doing when we try bagging?
  - A: Trying to average out variance. Reduce variance
  - Q: What do you think we're doing with boosting?

Boosting

- Boosting attempts to combine weak learners into strong learners
- How is this different than Bagging?
  - Q: What are we doing when we try bagging?
  - A: Trying to average out variance. Reduce variance
  - Q: What do you think we're doing with boosting?
  - A: We're combing weak learners, high bias, into a strong learner. Reduce bias

Boosting

- How do we do this?

Boosting

- How do we do this?
- General Idea: Train models on all data but at each iteration reweight data. Put more weight on data points incorrectly classified

Boosting

- How do we do this?
- General Idea: Train models on all data but at each iteration reweight data. Put more weight on data points incorrectly classified

Boosting:

- Initialize all data points with equal weight (1/n)

Boosting

- How do we do this?
- General Idea: Train models on all data but at each iteration reweight data. Put more weight on data points incorrectly classified

Boosting:

- Initialize all data points with equal weight (1/n)
- Loop:
  - Apply learner to weighted examples
  - Increase weights of misclassified examples

Boosting

- How do we do this?
- General Idea: Train models on all data but at each iteration reweight data. Put more weight on data points incorrectly classified

Boosting:

- Initialize all data points with equal weight (1/n)
- Loop:
  - Apply learner to weighted examples
  - Increase weights of misclassified examples
- Combine models by a weighted average

Boosting

- Instead of a single good classifier we get several classifiers that are bad on the entire dataset but better on certain areas of the dataset

Boosting

- Instead of a single good classifier we get several classifiers that are bad on the entire dataset but better on certain areas of the dataset
- Theory: It was proved that using boosting a weak learner can achieve zero training error

Boosting

- Instead of a single good classifier we get several classifiers that are bad on the entire dataset but better on certain areas of the dataset
- Theory: It was proved that using boosting a weak learner can achieve zero training error
- Is this dangerous? Why?

Boosting

- Instead of a single good classifier we get several classifiers that are bad on the entire dataset but better on certain areas of the dataset
- Theory: It was proved that using boosting a weak learner can achieve zero training error
- Is this dangerous? Why?
  - Unlike bagging we can overfit by adding more learners

Boosting

- Instead of a single good classifier we get several classifiers that are bad on the entire dataset but better on certain areas of the dataset
- Theory: It was proved that using boosting a weak learner can achieve zero training error
- Is this dangerous? Why?
  - Unlike bagging we can overfit by adding more learners
  - However, generally we can achieve good results

Boosting

- What you should know:
  - Combines weak learners, those performing just better than random (50%)
  - Results in a strong learner. Can achieve zero training error
  - Can overfit by adding more iterations but are very powerful in general
  - Many different algorithms: AdaBoost, Gradient Boosting, LogitBoost, etc…
  - Most popular AdaBoost, historically, and Gradient Boosting

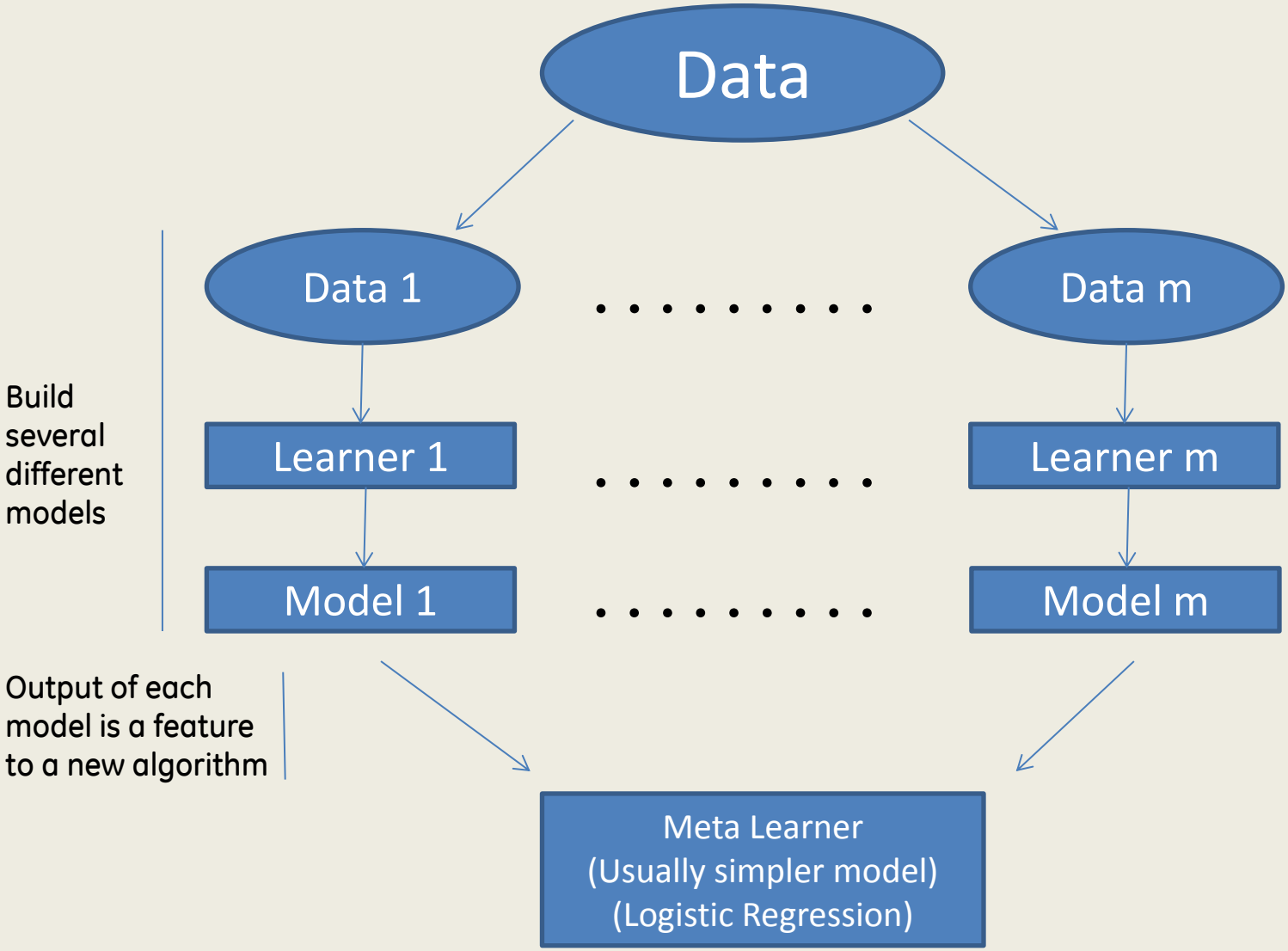
AdaBoost

- Input: Training set (S), learning algorithm, # rounds
- Initialize weights uniformly: w1(i) = 1/m
- Loop for number of rounds r:
  - Renormalize the weight
  - Train learner on weighted dataset
  - Measure the error rate, er, of the classifier on weighted S
  - If error > ½:
    - Exclude current learner and exit return all previous learners
  - Calculate learner performance
  - Reweight examples
- Return a weighted average

AdaBoost

- Input: Training set (S), learning algorithm, # rounds
- Initialize weights uniformly: w1(i) = 1/m
- Loop for number of rounds r:
  - Renormalize the weight: pr(i) = wr(i)/sum(wr(i))
  - Train learner on weighted dataset, hr
  - Measure the error rate, er, of the classifier on weighted S
    - er = sum(pr*I[hr(i) <> yi]
  - If error > ½:
    - Exclude current learner and exit return all previous learners, r-1
  - Calculate learner performance: br = er/(1-er)
  - Reweight examples: w(r+1)(i) = wr(i)b_r^(1-I[hr(xi)<> yi])
- Return a weighted average: sum(log(1/b_r)*hr(x))

Stacked Generalization



- Questions?