

Introduction to Machine Learning

Day 7: Boosting

Joanna Bieri DATA301

Get out your handwritten notes!

One At A Time

ASK

Last week the trees all worked at once,
and then voted.

Today they go one at a time.

So what does the next tree work on?

Hint. What has nobody explained yet?

One At A Time

ANSWER

- **The leftovers.** Whatever the model still gets wrong, which we call the residual.
- They are **added together**, not voted on, each one a small fix.
- A forest tree ignores the others. A boosted tree fixes them.

One Small Correction

PREDICT

Your model says a house is worth 3.0. It really sold for 4.2.

The next tree is trained on what is left over, and it predicts 1.2. Your learning rate is 0.1.

What does the model say after that tree is added?

Hint. First: how wrong were we? Then: we only add a tenth of what the new tree says.

One Small Correction

ANSWER

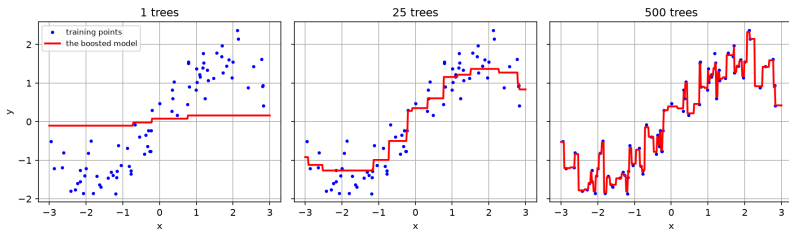
$$\text{leftover:} \quad 4.2 - 3.0 \quad = 1.2$$

$$\text{new guess:} \quad 3.0 + 0.1 * 1.2 \quad = 3.12$$

It moves a little. That is what the learning rate does, and it is why boosting needs a lot of trees.

Watching It Build

ASK



Which one of these would you ship? What is the last one doing?

Hint. The dots are a smooth curve with random noise added on top.

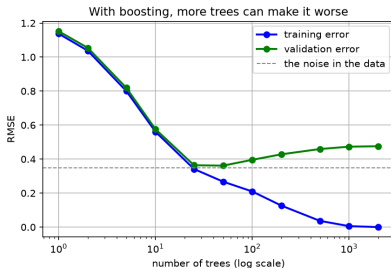
Watching It Build

ANSWER

- **The middle one.** It found the curve and ignored the wiggles.
- The last one traces individual points, noise included.
- The first is two flat steps. One tree of depth 2 cannot do more.

More Is Not Better

ASK



Last week more trees never hurt. Why does this turn around?

Hint. Once the real pattern is used up, what is left to fix?

More Is Not Better

ANSWER

- Each tree fixes leftovers. Once the pattern is gone, the leftovers **are** the noise.
- The green line bottoms out near 50 trees and gets worse from there. The blue line goes to zero.
- A forest averages trees that never see each other, so it cannot do this to itself.

Small Steps Or Big Steps

ASK

You cut the learning rate to a tenth
of what it was.

What else has to change?

Hint. Each tree is only added a little at a time. So how far do you get?

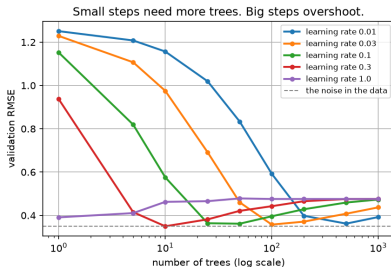
Small Steps Or Big Steps

ANSWER

- **About ten times more trees.** In the notes, 50 became 500.
- Those two settings are **one decision**: lower one, raise the other.
- Every rate we tried landed at about the same error. They just took different numbers of steps.

When The Steps Are Too Big

ASK



One of these never gets to a low error. Which one, and why?

Hint. What happens if you add the whole correction instead of a slice of it?

When The Steps Are Too Big

ANSWER

- The learning rate of **1.0**. It overshoots, then spends later trees undoing the overshoot.
- Its best result came from a **single tree**, which should feel strange.
- Big steps struggle to converge.

It Stopped By Itself

ASK

You offer XGBoost two thousand trees.

It uses a couple hundred and quits.

What made it stop?

Hint. It was watching something while it trained.

It Stopped By Itself

ANSWER

- It watched the **validation** score and stopped when that stalled.
- `early_stopping_rounds=50` is patience: fifty trees in a row with no gain, then stop and keep the best.
- We set the limit at 2000 and it used 193. We never had to guess the number.

What It Watches

ASK

Early stopping has to watch some data.

Why not let it watch the test set?

Hint. Is "how many trees" a choice?

What It Watches

ANSWER

- Because it **is** a choice, and choices get made on validation data.
- A test set that picked the tree count helped build the model, so its score is no longer honest.
- This is the same rule as Day 4 and Day 5.

Breaking The Tie

ASK

Two models, tied.

The test set would settle it today.

Do we open it?

Breaking The Tie

ANSWER

- **No.** Score every new model on it and it becomes a second validation set.
- Break the tie on training time or simplicity.

Questions?

Questions?