Boosting, Gradient Boosting and AdaBoost

Introduction to Boosting

CS1090A Introduction to Data Science

Pavlos Protopapas, **Natesh Pillai**, Chris Gumb

Jingfei Huang
Moshi Park, Sichuan, China

Outline

- Introduction to Boosting
- Gradient Boosting
- Mathematical Formulation - Gradient Boosting

Outline

- **Introduction to Boosting**
- Gradient Boosting
- Mathematical Formulation - Gradient Boosting

Boosting

Anthony Goldbloom gives you the secret to winning Kaggle competitions

January 13, 2016 Andrew Fogg Big Data

Kaggle has become the premier Data Science competition where the best and the brightest turn out in droves – Kaggle has more than 400,000 users – to try and claim the glory. With so many Data Scientists vying to win each competition (around 100,000 entries/month), prospective entrants can use all the tips they can get.

And who better than Kaggle CEO and Founder, Anthony Goldbloom, to dish out that advice? We caught up with him at Extract SF 2015 in October to pick his brain about how best to approach a Kaggle competition.

Anthony Goldbloom gives you the secret to winning Kaggle competitions

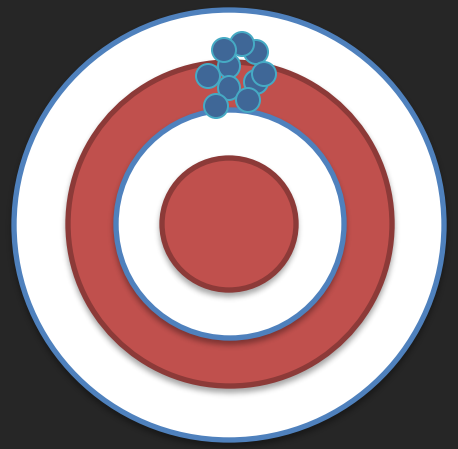
As long as Kaggle has been around, Anthony says, it has almost always been ensembles of decision trees that have won competitions.

It used to be random forest that was the big winner, but over the last six months a new algorithm called XGboost has cropped up, and it's winning practically every competition in the structured data category.

Recap: Decision Trees Issues

- Shallow trees:

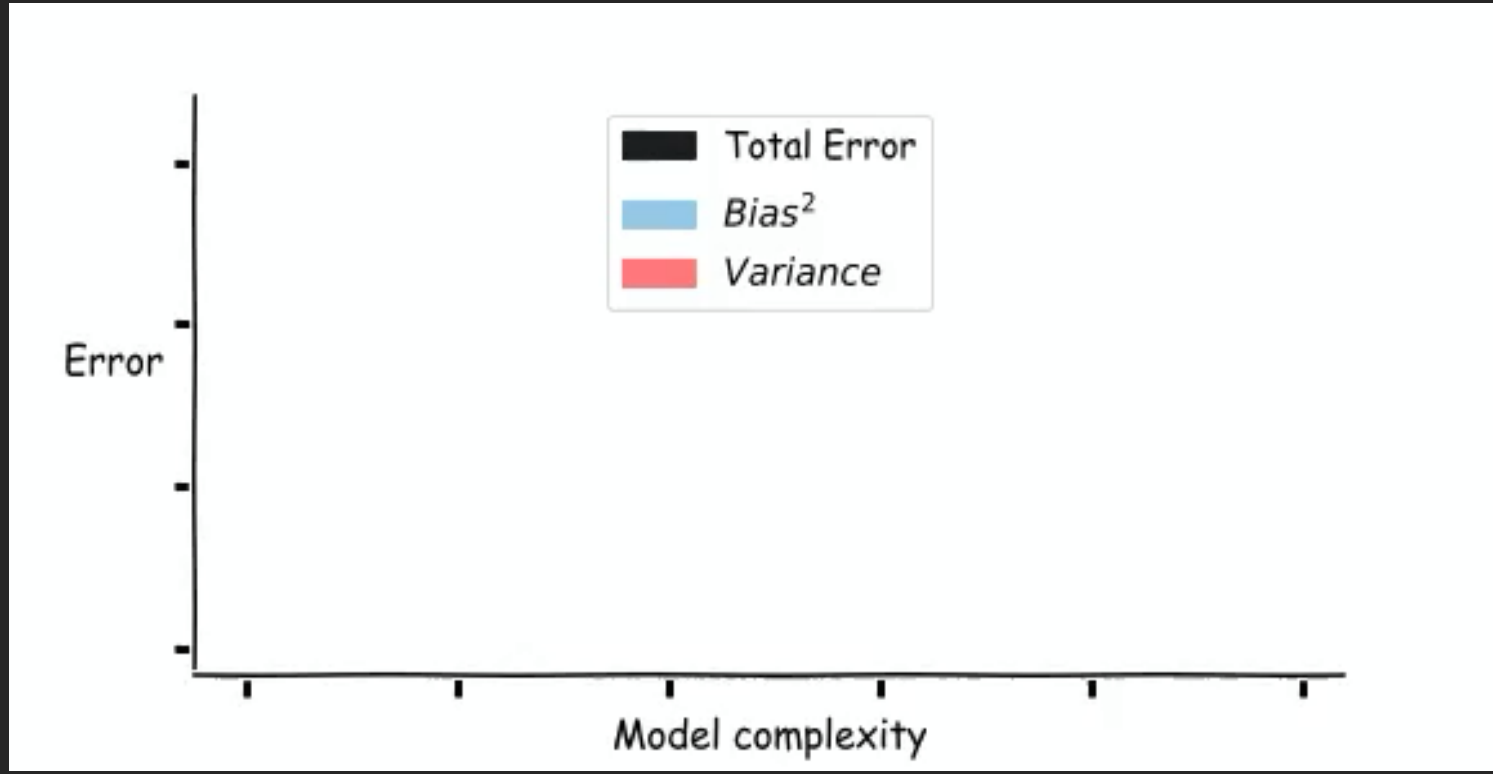
  Shallow trees (with very few leaves) suffer from high bias and do not train well.



- Deep Trees:

  Deep trees (with large number of nodes and leaves) have low bias but suffer from high variance leading to very low generalization error.

Recap: Decision Trees - Bias-Variance Trade-off



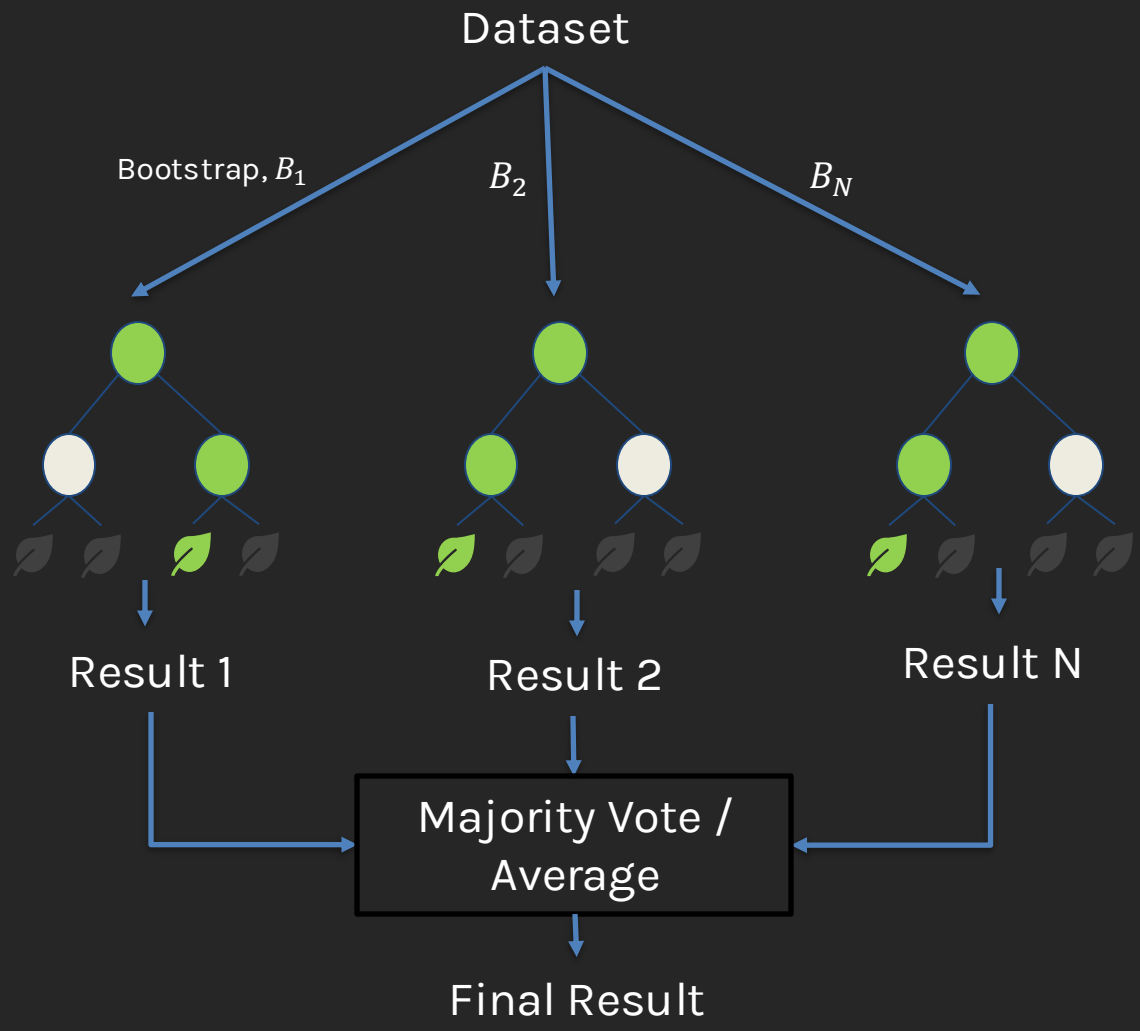
Random Forest Issues

- **Variance:**

  Although variance reduction is better in RF than bagging, the generalization error is still high.

- **Inference Speed:**

  Large number of trees can make the algorithm very slow and ineffective for real-time predictions.



Motivation for Boosting

Question: Could we address the shortcomings of single decision tree models in some other way?

For example, rather than performing **variance reduction** on complex trees, can we **decrease the bias** of simple trees - make them more expressive?

Can we learn from our mistakes?

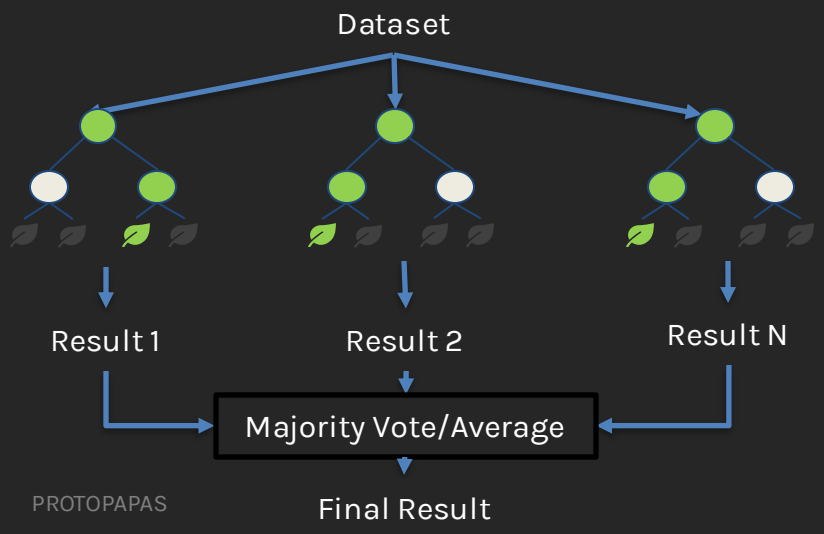
A solution to this problem, making an expressive model from simple trees, is another class of ensemble methods called **boosting**.

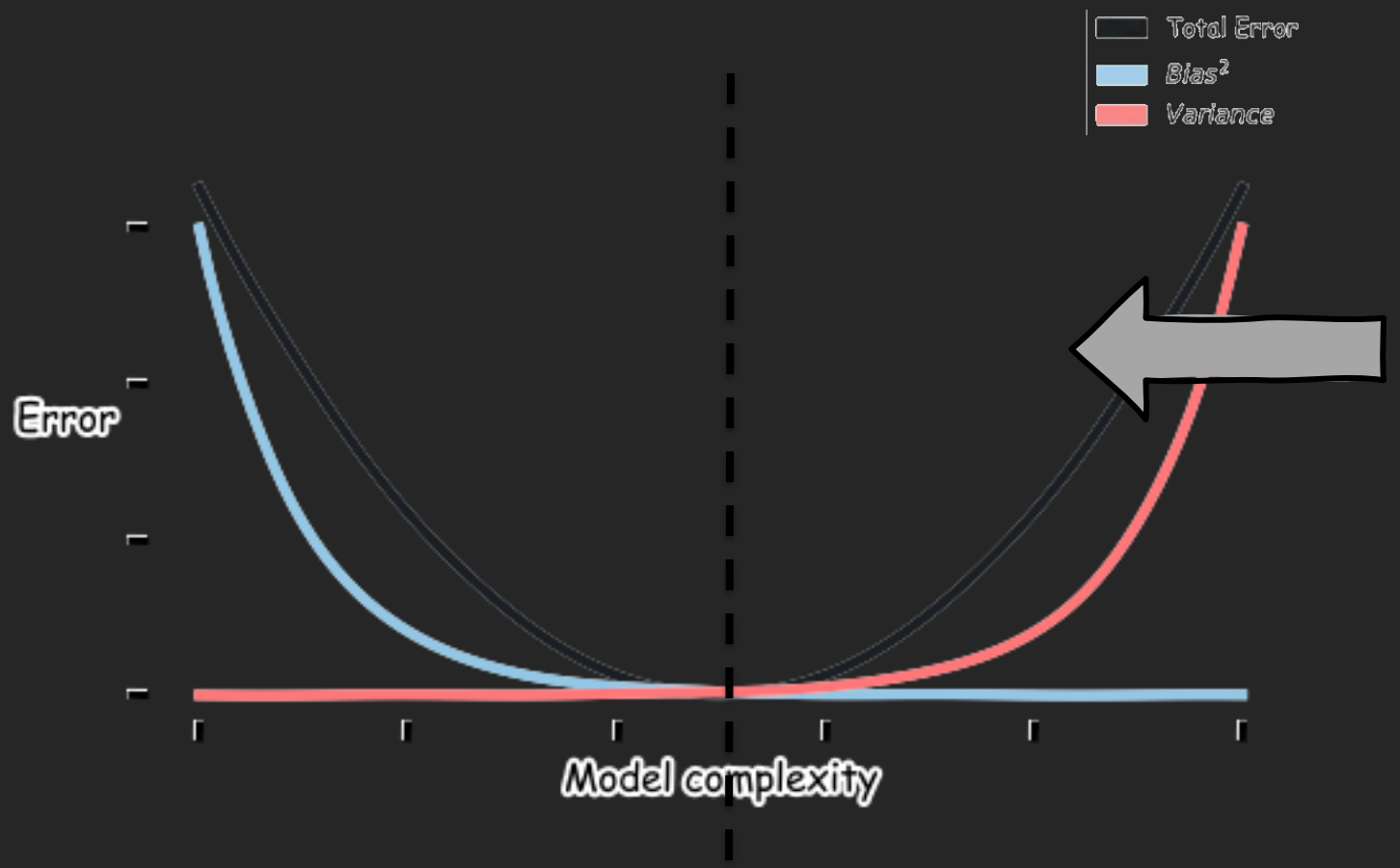
Random Forest – The only solution ?

OPTION #1

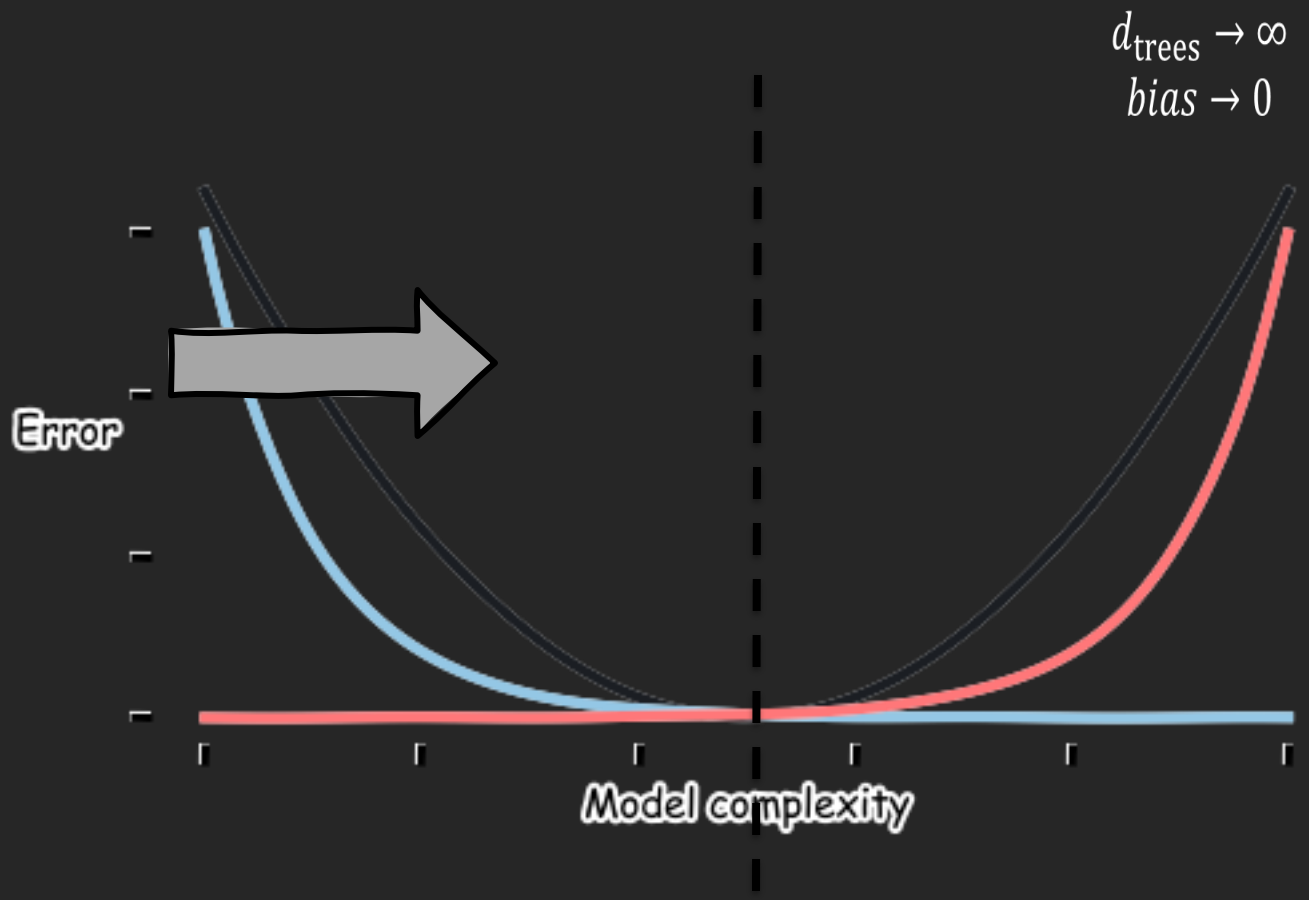
Reduce variance

$$d_{\text{trees}} \to \infty$$
$$var \to 0$$





Random Forest – The only solution ?



OPTION #2

Reduce bias

$$d_{\text{trees}} \to \infty$$
$$bias \to 0$$

?

Boosting

NEW IDEA 💡

- **Boosting** methods are general algorithms which combine several "**weak learners**" to produce a strong rule.
- The first implementation of Boosting was '**Adaboost**' invented by Robert Schapire and Yoav Freund in 1996.
- Boosting algorithms are fast, easy to compute and very accurate and are the de-facto optimization tree algorithms.

Rob Schapire & Yoav Freund

How does boosting work?

FINAL EXAM

TOPIC: BOOSTING DATE: DEC-11-2024

Q1:

Q2:

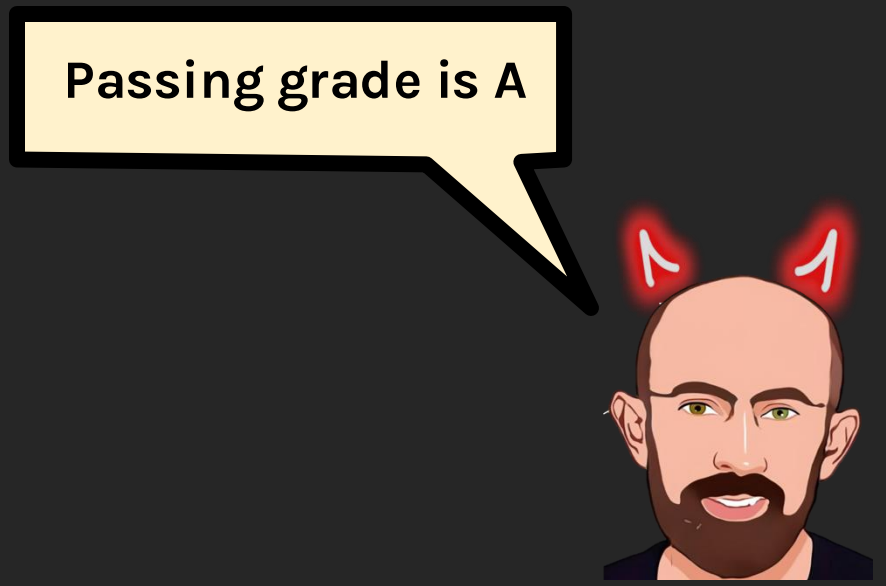
Q3:

...

Q10:

final score

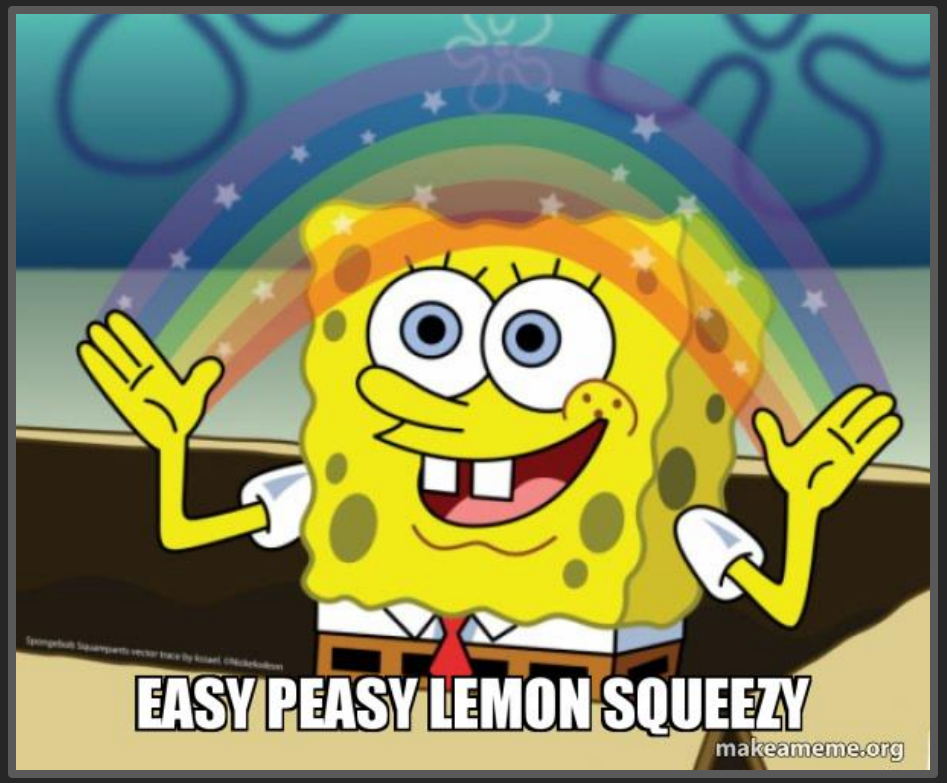
_ _ _ _ _ _ _ _



How does boosting work?

OPTION #1

1. Steal a time machine.
2. Go back to 1996 and meet Rob Schapire and Yoav Freund.
3. Befriend them by giving them stock trading tips from the future.
4. Follow their work for at least a decade to understand everything about boosting.
5. Return to the present and nail the test.
6. Repeat for another test.

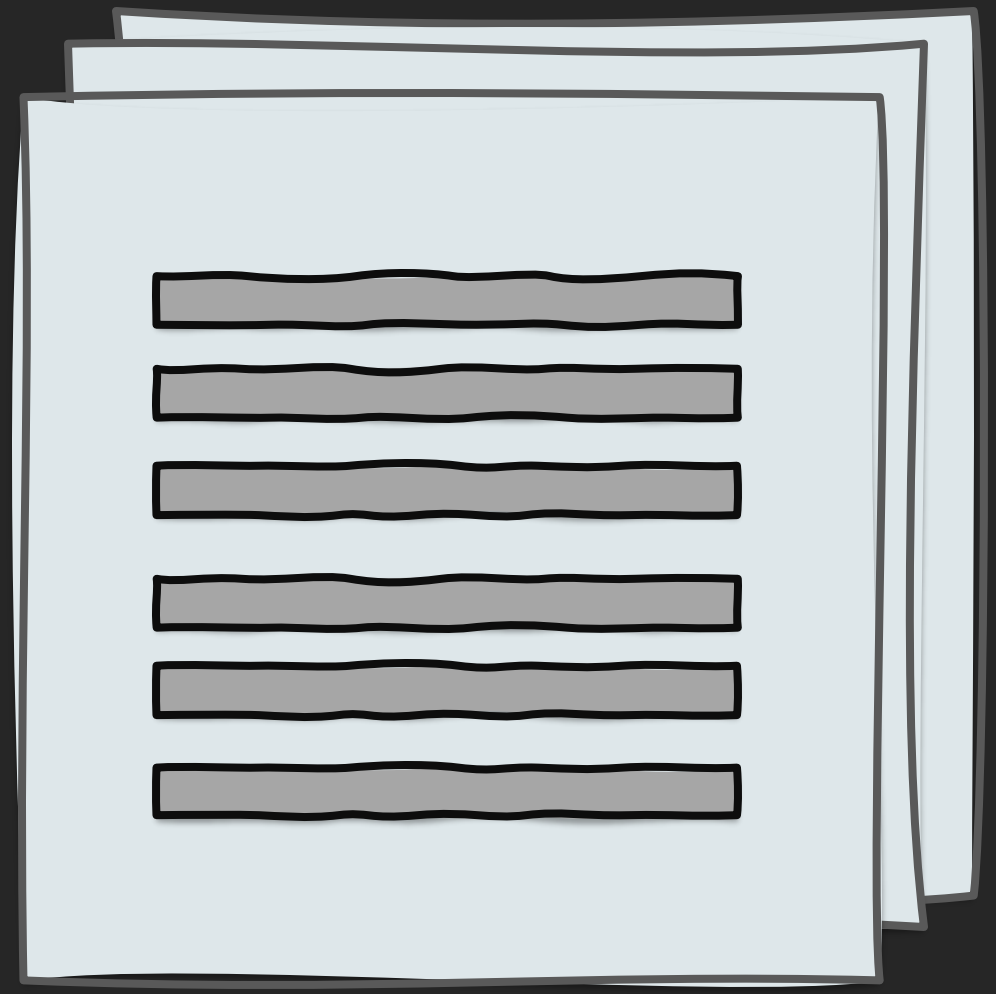


How does boosting work?

OPTION #2

STEP #1:

Go to the library and get previous year question papers.

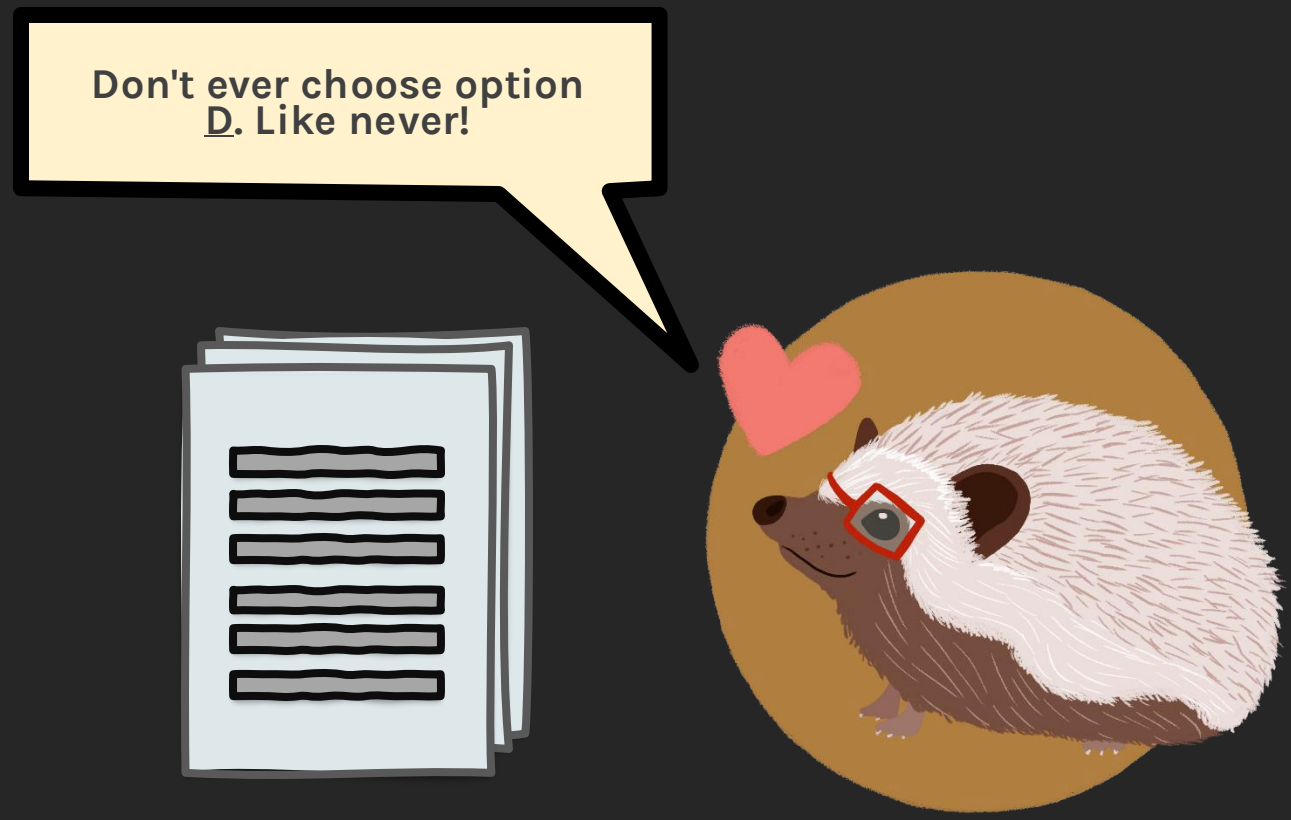


How does boosting work?

OPTION #2

STEP #2:

Find a helpful student and ask them to give you a "rule of thumb" to get at least some answers right.



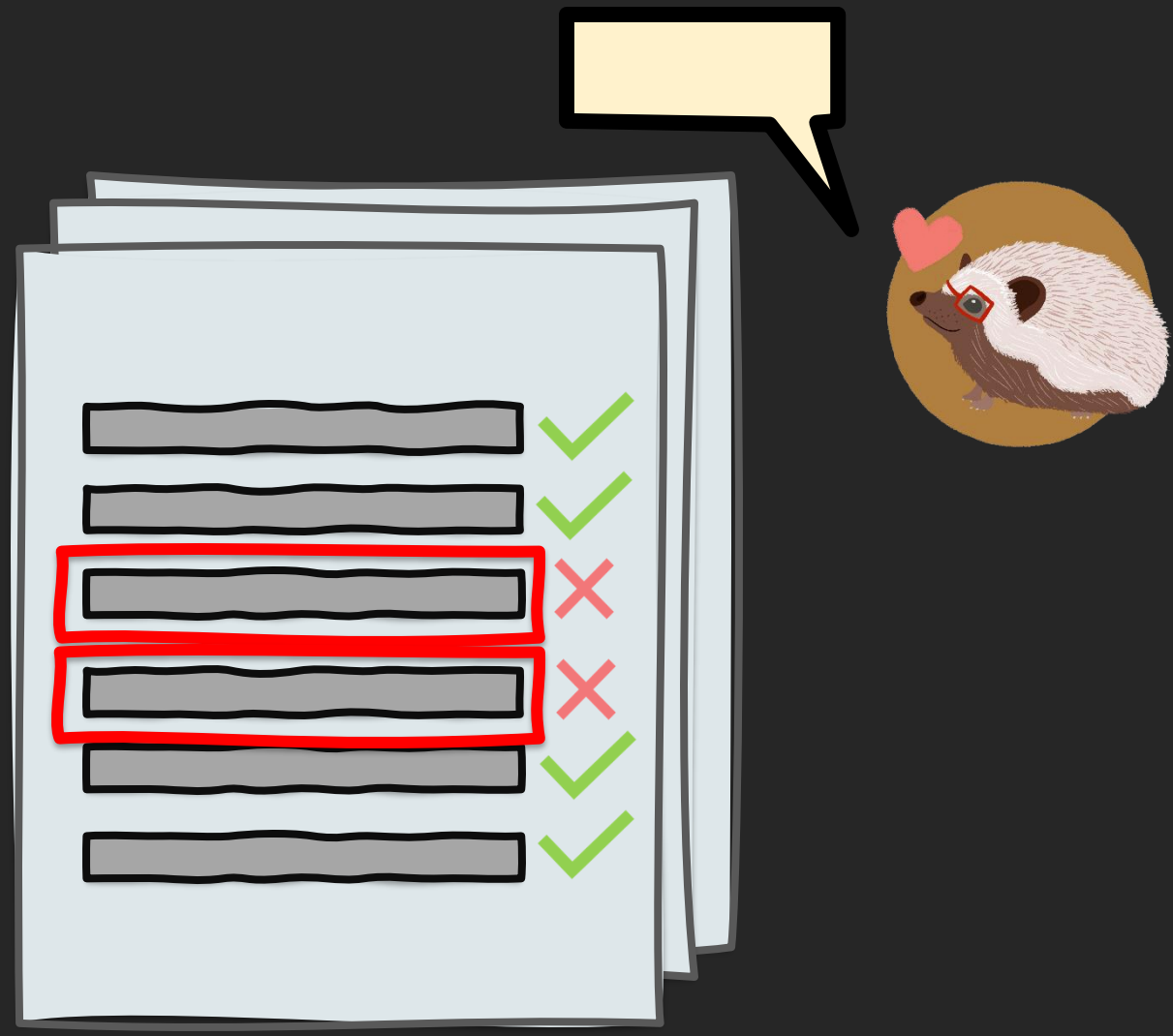
How does boosting work?

OPTION #2

Does the “rule” work?

Test out the rule.

It worked 60% of the time.
Not bad!!



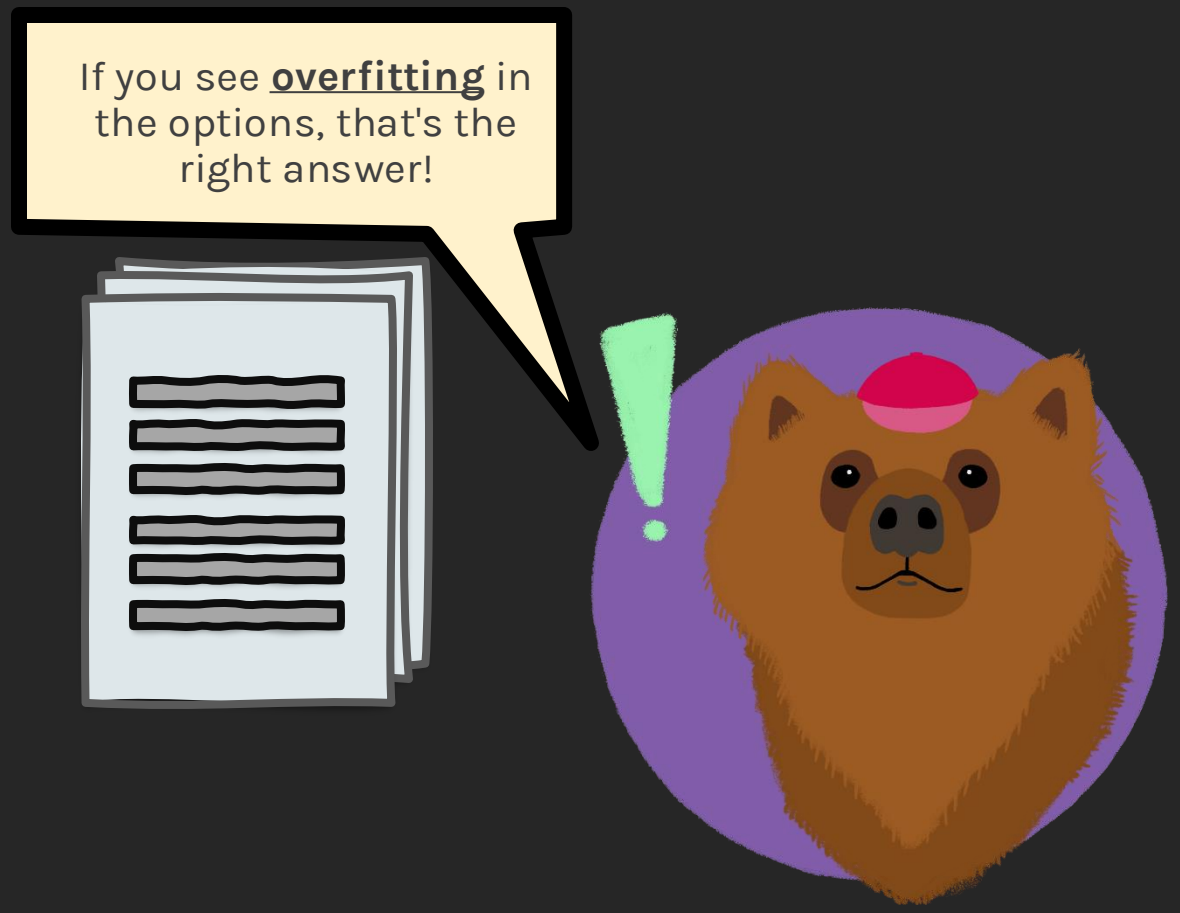
How does boosting work?

OPTION #2

STEP #3:

Find a TA and ask them to also give you a "rule of thumb" to get at least some answers right.

Make sure to focus on the ones you got wrong before!



How does boosting work?

OPTION #2

Does the “rule” work?

Test out the new rule.

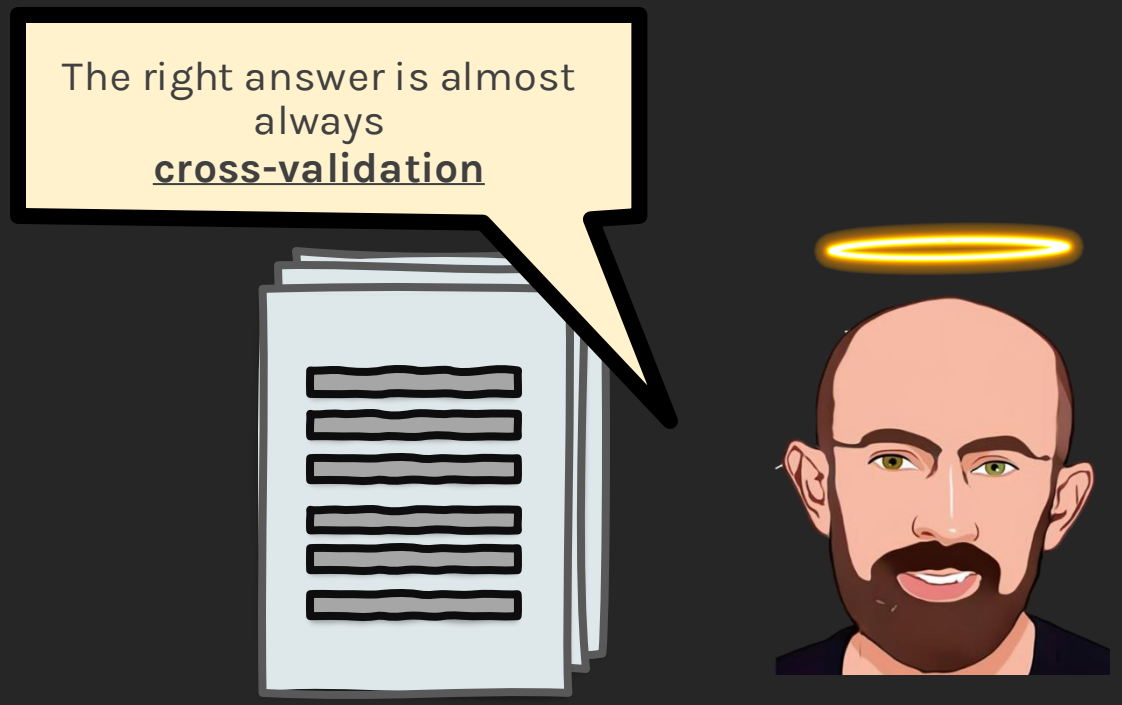
It works well on difficult problems!
But a few problems persist.

How does boosting work?

OPTION #2

STEP #4:

Call your favorite professor and focus on the ones you got wrong before!



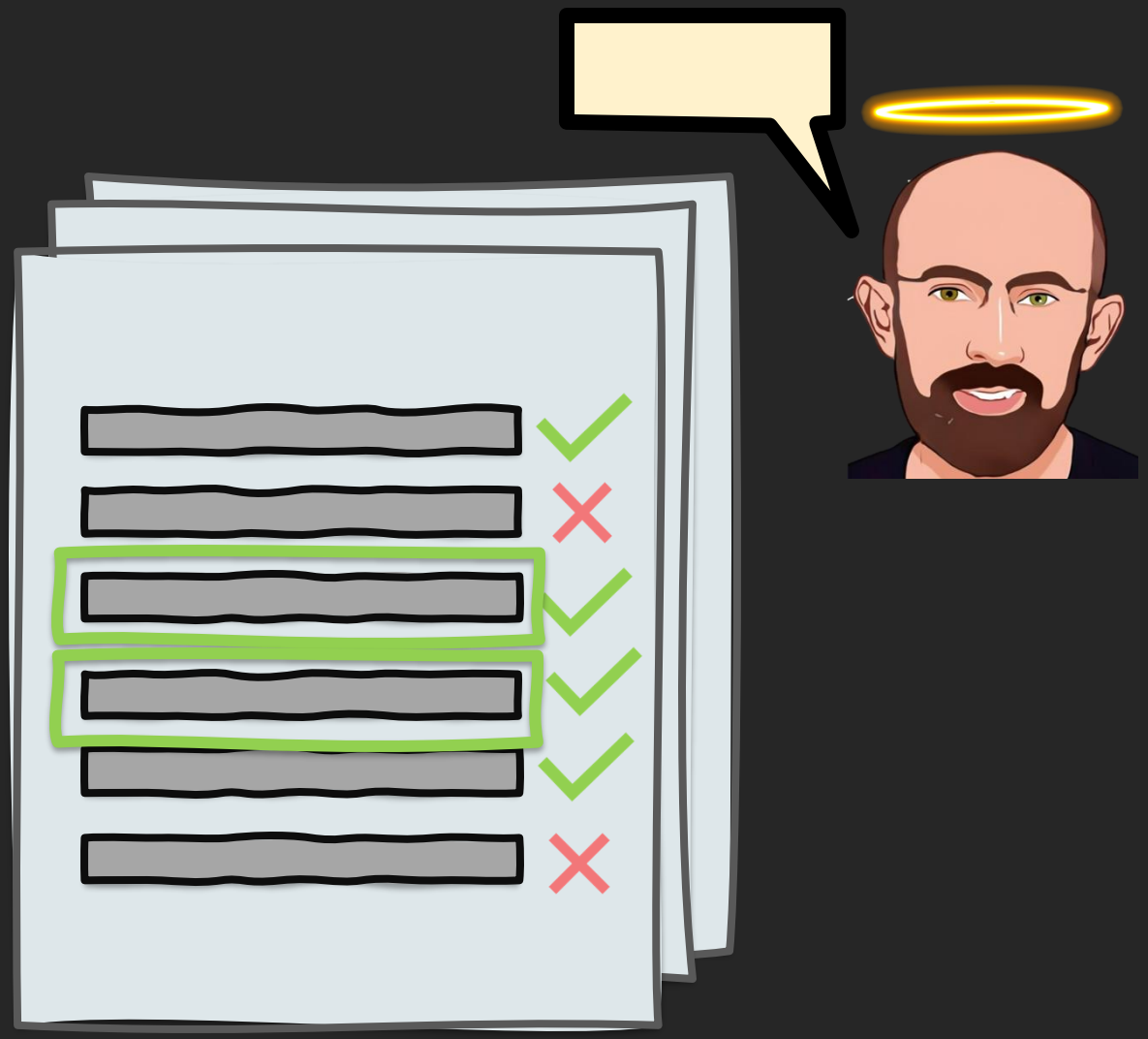
How does boosting work?

OPTION #2

Does the “rule” work?

Test out the new rule.

The new rule works well on the difficult problems!

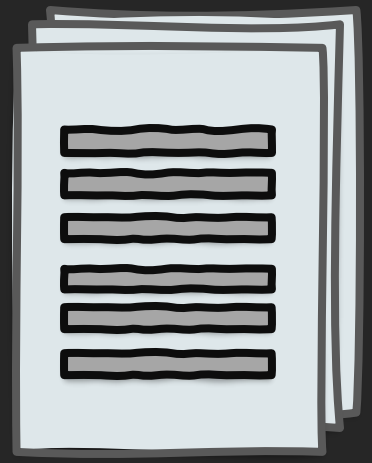


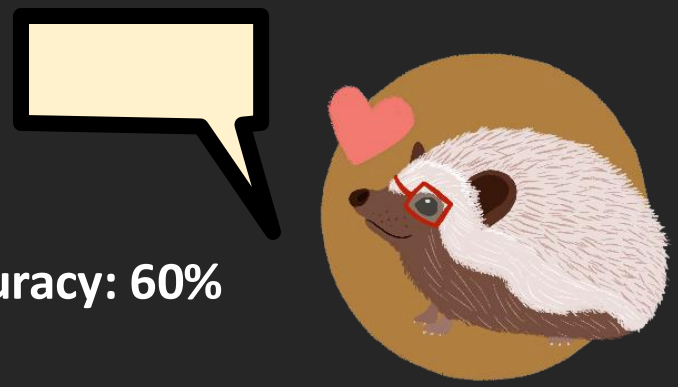
How does boosting work?

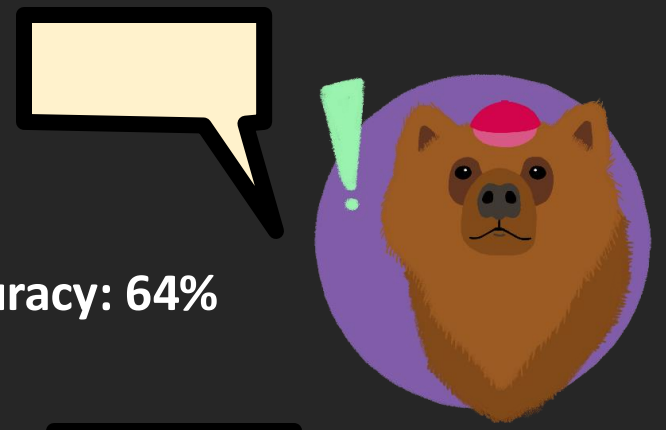
OPTION #2

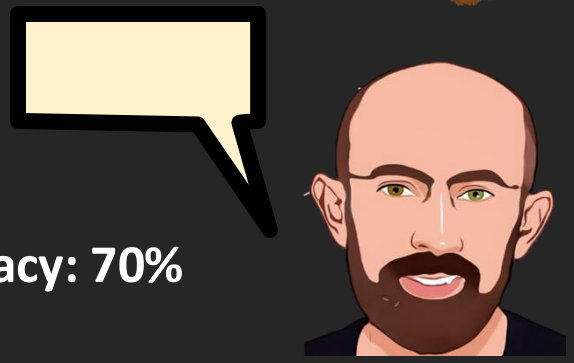
STEP #5:

Combine the rules, but pay more attention to the ones that were more often right.





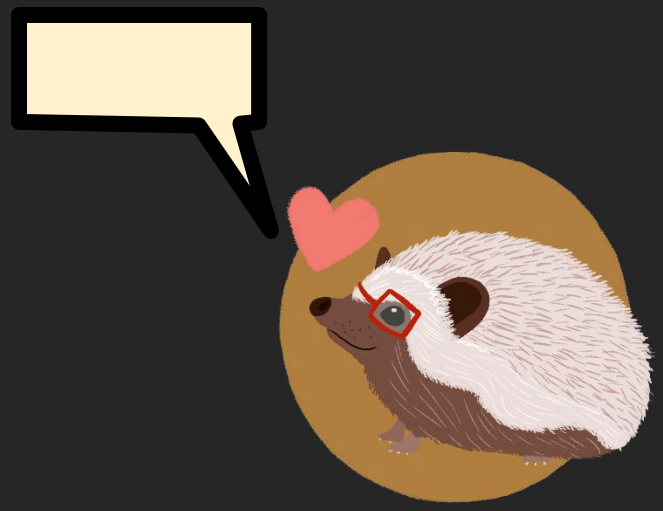




How does boosting work?

OPTION #2

Accuracy: 60%



Accuracy: 64%

Accuracy: 70%

$$Strategy = \alpha * Rule_1 + \beta * Rule_2 + \gamma * Rule_3$$

How does boosting work?

OPTION #2

FINAL STEP:

Take the test with these approximate rules, weighted by how well each rule performed.

A+

But how does it work?