Elements of Machine Learning & Data Science

Winter semester 2023/24

Automated Machine Learning (1)

Prof. Holger Hoos

Key questions:

- How good is an ML model?
- How good could an ML model be?

Key questions:

- **How good is an ML model?**
  - Is it “fit for use” (i.e., good enough for deployment)?
  - What are its strengths and weaknesses?
  - Might anything have gone wrong during training?

Key questions:

- **How good is an ML model?**
  - How do we assess whether it is “fit for use” (i.e., good enough for deployment)?
  - How do we assess its strengths and weaknesses?
  - How do we detect if anything has gone wrong during training?

Key questions:

- **How good could an ML model be?**
  - Are we using the best possible ML method / model?
  - Have we configured and trained it in the best possible way?
  - Can we further improve performance?

Key questions:

- **How good could an ML model be?**
  - How can we ensure we are using a good ML method / model?
  - How can we configure and train it for optimised performance?
  - How can we further improve performance?

High-level learning goals:

Be able to …

- answer these key questions in a technical manner;
- recognise weaknesses in the empirical performance of ML models using standard tools and methods;
- explain these analysis tools and methods at a technical level;
- use standard tools and methods for selecting models and optimising their hyperparameters;
- explain these AutoML tools and methods at a technical level.

TPS Exercise

You want to solve a supervised classification problem.

Question:

How to make sure you to construct the best possible classifier, i.e., which methods and techniques should be applied?

Automated Machine Learning

Key idea:

Automate the decisions that have to be made when constructing effective machine learning models / pipelines

Methods/approaches:

- model selection -> today (Holger Hoos)
- hyper-parameter optimisation (HPO) -> next Tuesday (Anja Jancovic)
- neural architecture search (NAS) -> next Friday (Marie Anastacio)
- combined algorithm selection and hyperparameter optimization (CASH), automating data science -> Tuesday, 30 January (Holger Hoos)

All four classroom sessions: Inverted classroom approach

Key idea:

- You learn relevant concepts and approaches via assigned pre-class homework (reading, answering questions, …)
- We use classroom time to clarify, reinforce, expand, discuss

Advantages:

- You have more control over your individual learning process (but also more responsibility)
- We use classroom time more effectively
- Classes should be less boring and more fun

Note: For this to work, it is essential that you do homework, participate and take notes.

Preparation for today:

Read the following research paper:

Oded Maron and Andrew Moore: Hoeffding Races: Accelerating Model Selection Search for Classification and Function Approximation. Advances in Neural Information Processing Systems 6 (NIPS 1993): 59-66, 1993.
(The paper is available online at https://proceedings.neurips.cc.)

Focus on the following questions (which will be further explored in TPS exercises in class):

(1) What is the fundamental problem when using cross-validation
(or performance on a validation set) to select between different ML models?
(2) What is the key idea behind Hoeffding Races and how does it address the problem
identified in (1)?
(3) What is the role of the parameters Δ and δ, respectively?

Bring your answers to these questions (which can be in the form of bullet points) to class; they will be the basis for TSP exercises).

NB: Full understanding of the proof in Section 3 is desirable but not essential.

AI StudyBuddy

TPS Exercise (T part = done as preparation for class)

You have read the research paper by Maron & Moore about Hoeffding Races for model selection.

Question:

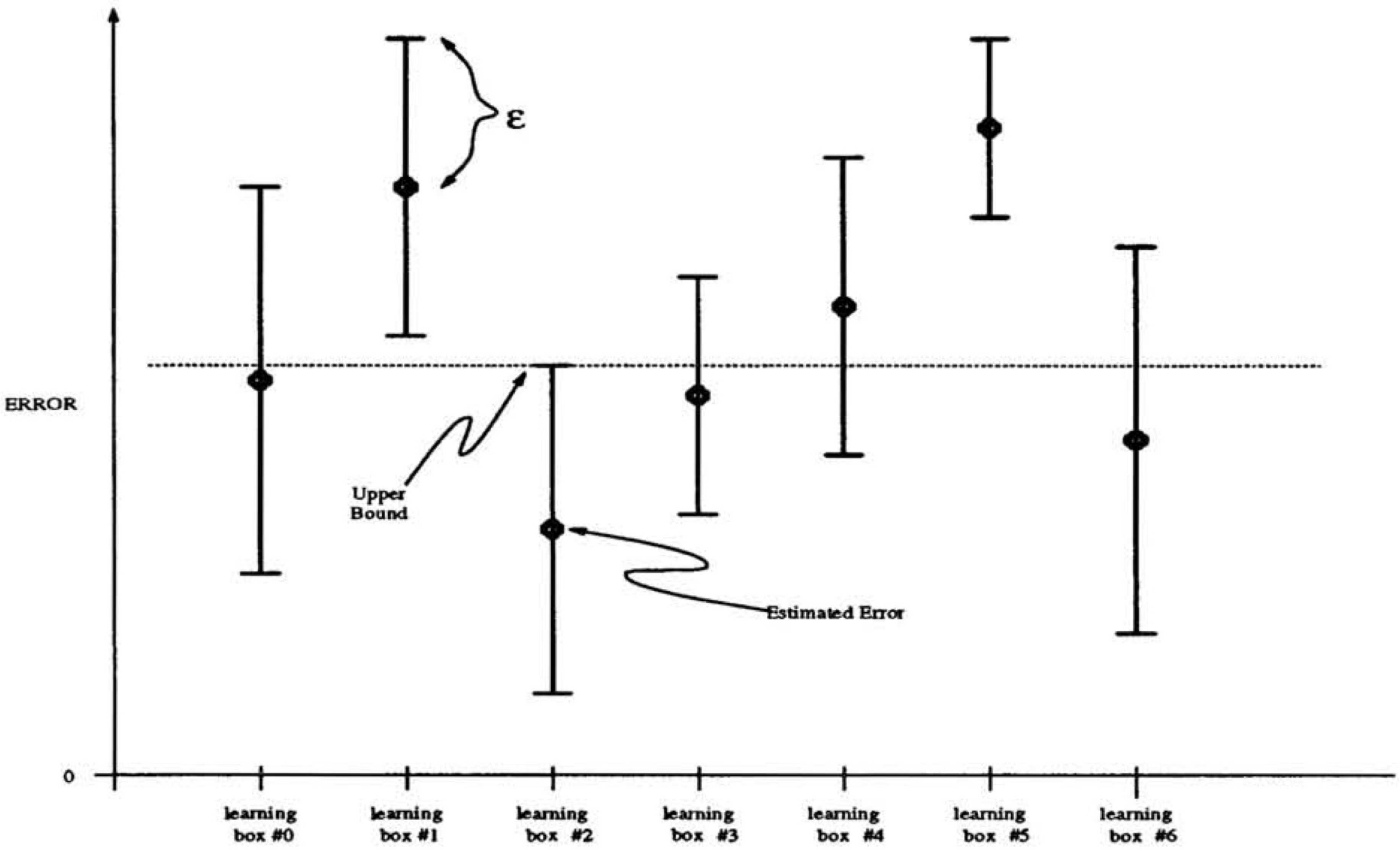
What is the fundamental problem when using cross-validation (or performance on a validation set) to select between different ML models?

TPS Exercise (T part = done as preparation for class)

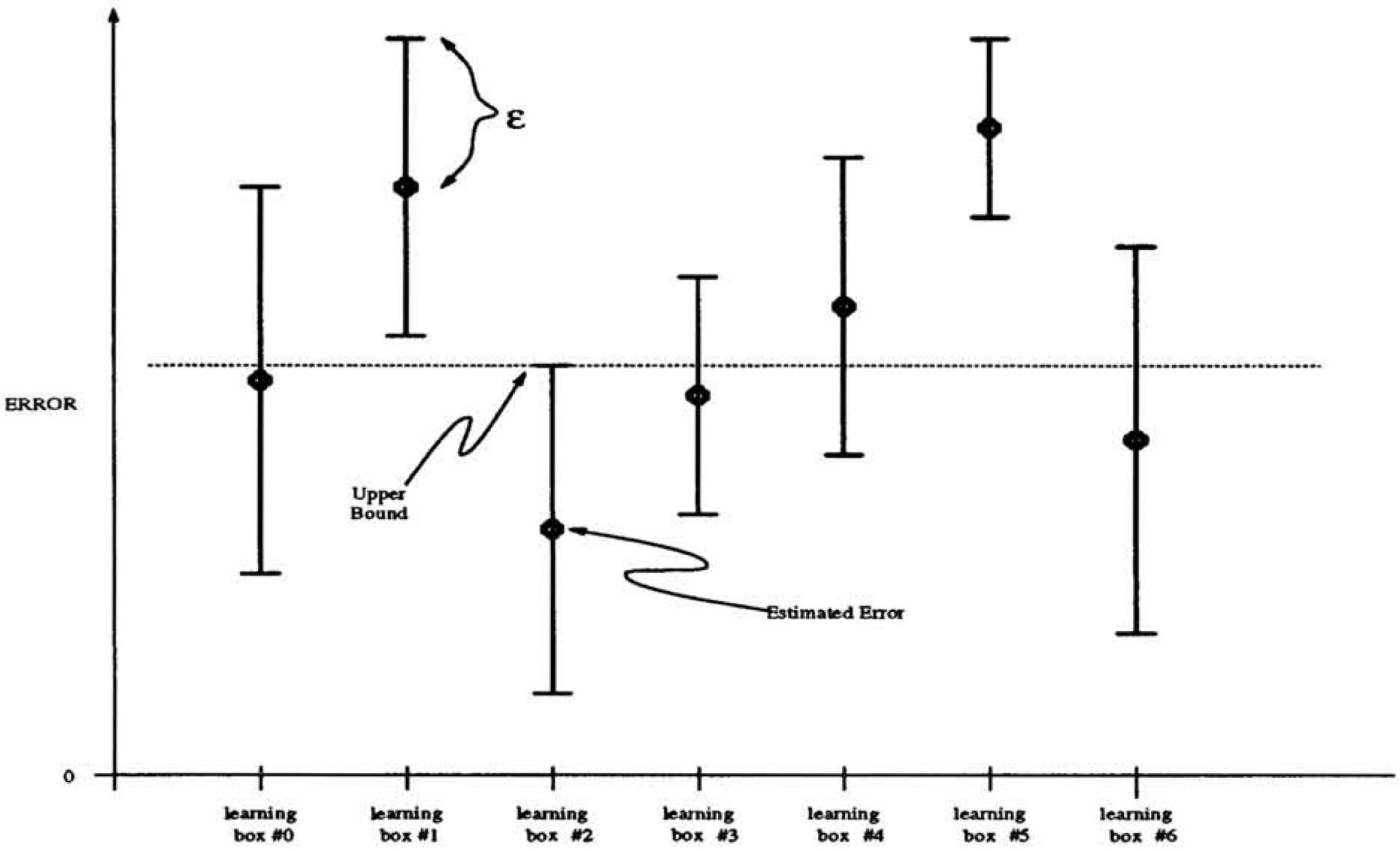
You have read the research paper by Maron & Moore about Hoeffding Races for model selection.

Question:

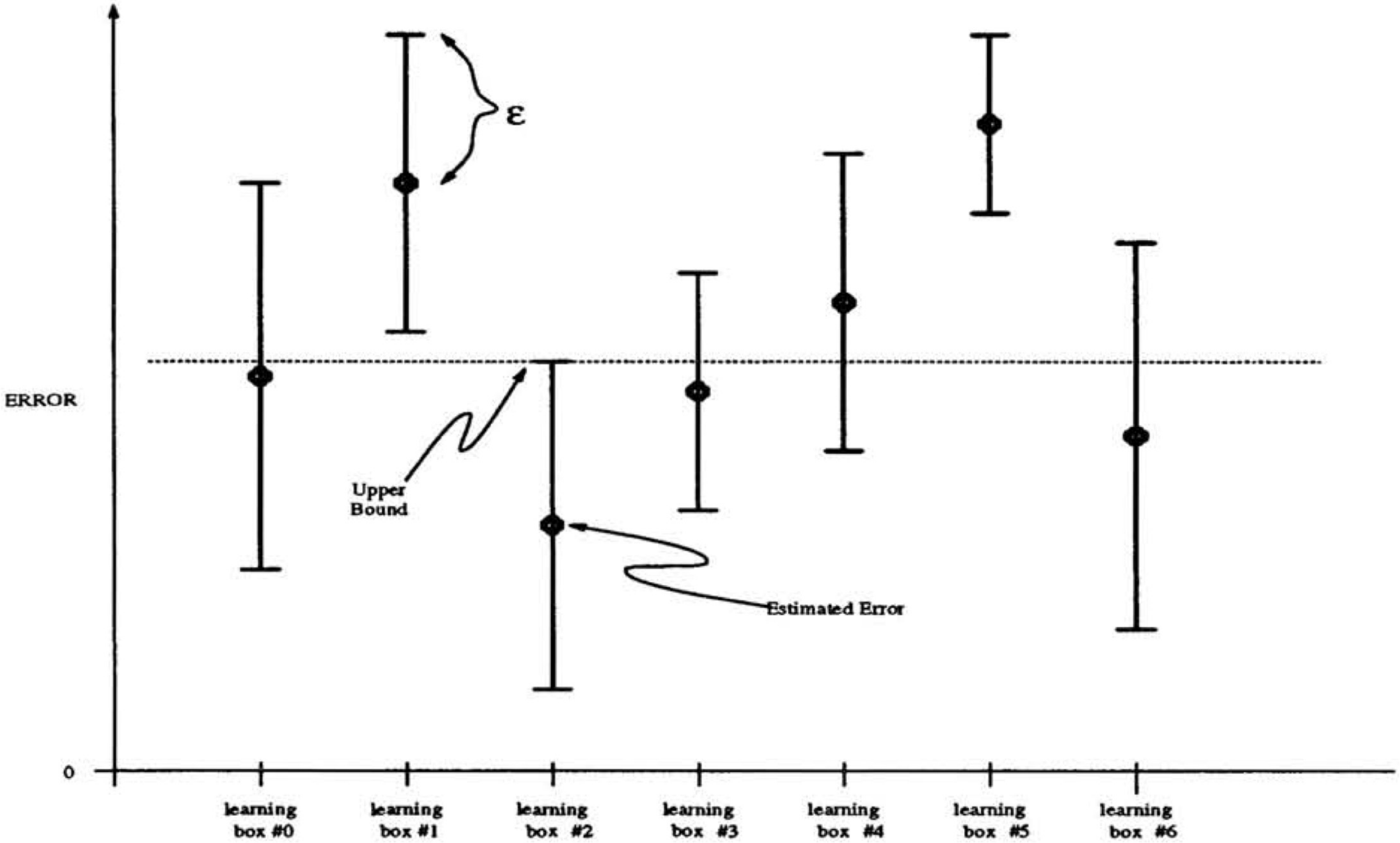
What is the key idea behind Hoeffding Races?



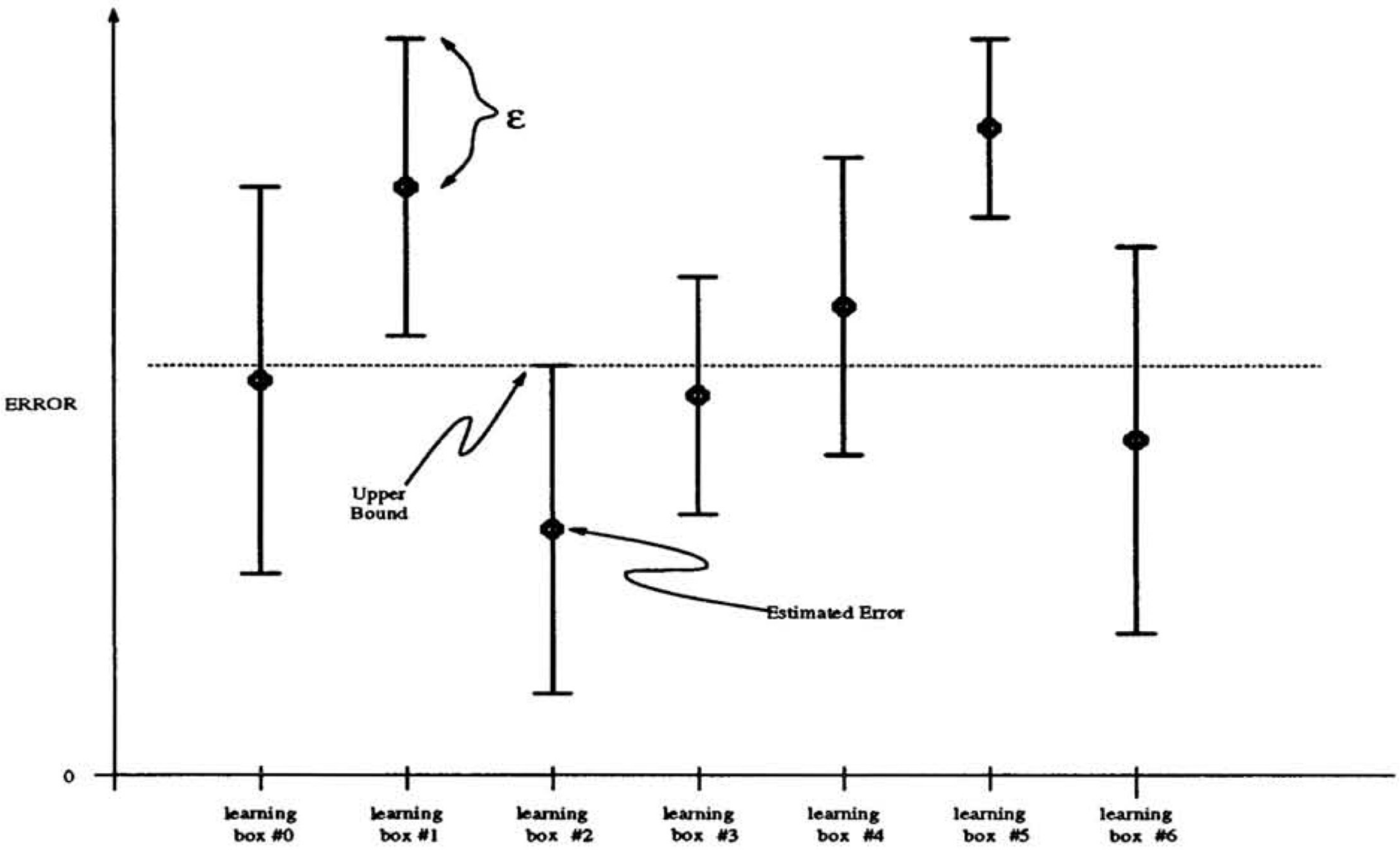
Which models (“learning boxes”) can we safely eliminate?



Which models ("learning boxes") can we safely eliminate? #1, #5



What happens as more & more inputs are being tested?



What happens as more & more inputs are being tested? Error bars shrink

At each point in the algorithm, we $\boxed{\text{randomly select}}$ a point from the test set. We compute the error at that point for all learning boxes, and update each learning box's estimate of its own total error rate. In addition, we use Hoeffding's bound to calculate how close the current estimate is to the true error for each learning box. We then eliminate those learning boxes whose best possible error (their lower bound) is still greater than the worst error of the best learning box (its upper bound); see Figure 2. The intervals get smaller as more points are tested, thereby "racing" the good learning boxes, and eliminating the bad ones.

Why is it important to select (uniformly) at random?

We repeat the algorithm until we are left with just one learning box, or until we run out of points. The algorithm can also be stopped once ϵ has reached a certain threshhold. The algorithm returns a set of learning boxes whose error rates are insignificantly (to within ϵ) different after N test points.

What should we do if we run out of points with more than one model (“learning box”) left?

TPS Exercise (T part = done as preparation for class)

You have read the research paper by Maron & Moore about Hoeffding Races for model selection.

Question:

What is the role of the parameters Δ and δ, respectively?

δ: upper bound on P (| true error – estimated error | > ε)
for *one* algorithm, during *one* stage (iteration) of the race

i.e., lower δ -> lower probability that models get eliminated incorrectly

Δ: overall probability that best model gets eliminated incorrectly

$\Delta = \delta$ * #models * |test set|

Note: Choose Δ, derive δ from this, then ε (used within HR)

AI StudyBuddy

TPS Exercise

Question:

How could Hoeffding Races be further improved?

TPS Exercise

Question:

How could Hoeffding Races be further improved?

- Better training data -> better models to start with
- Predict model performance to pre-select models
- Use hyper-parameter optimisation to improve models

Key concepts covered today:

- AutoML
- model selection
- brute-force search
- Hoeffding Races

Preparation for class on Tuesday, 23 January 2024 (mandatory):

Read the following research paper:

James Bergstra and Yoshua Bengio: Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research (JAIR): 281-305, 2012.
(The paper is available online at https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf)

Focus on the following questions (which will be further explored in TPS exercises in class):

(1) What is the fundamental problem when using grid (and manual) search for hyper-parameter optimization? Explain what a response surface is in simple terms.
(2) What is low effective dimensionality and how does it allow for success of random search to address the problem identified in (1)?
(3) What is the key strength of the sequential combination of manual and grid search?

Bring your answers to these questions (which can be in the form of bullet points) to class; they will be the basis for TPS exercises).

NB: Full understanding of the theoretical sections (e.g., 2.1) is desirable but not essential.