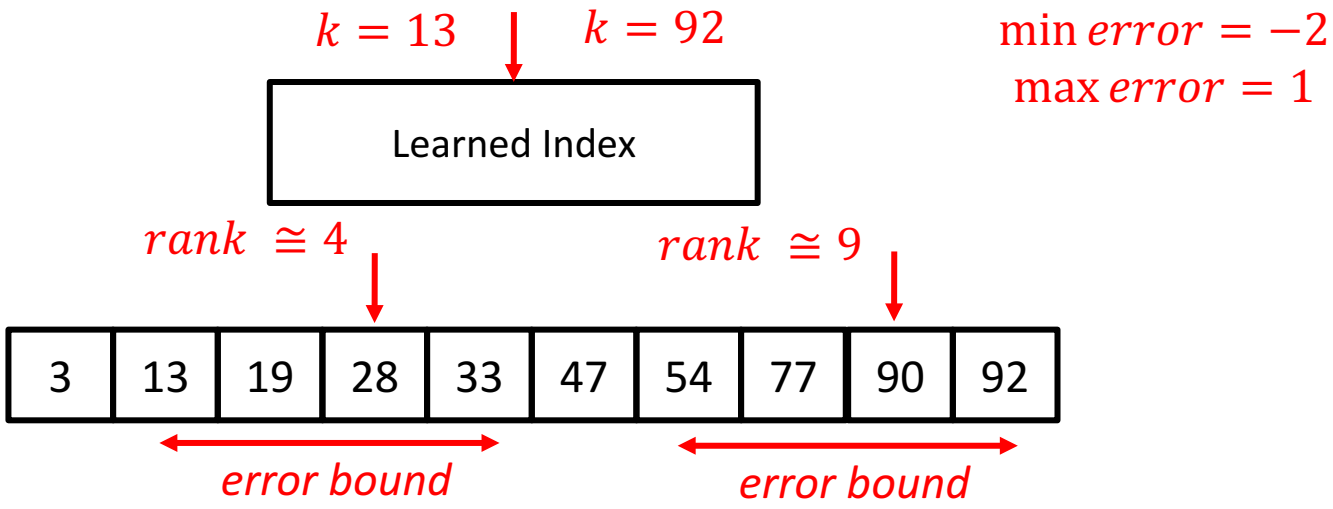
Lab 1 – Learned Index Structures

- Build a 2-stage recursive model index
- Simulate with python, numpy and Tensorflow
- Both synthetic and real-work workloads will be provided

Problem Formulation

- Input
  - a key *k*
- Output
  - the *estimated rank* of *k* among all the keys



Recursive Model Index (RMI)

- If the error is too large, the last mile search would be very costly
- Key observation
  - reducing error to 100 from 100M is hard
  - reducing error to 10K from 100M is much easier; similarly, reducing error to 100 from 10K is much easier

Recursive Model Index (RMI)

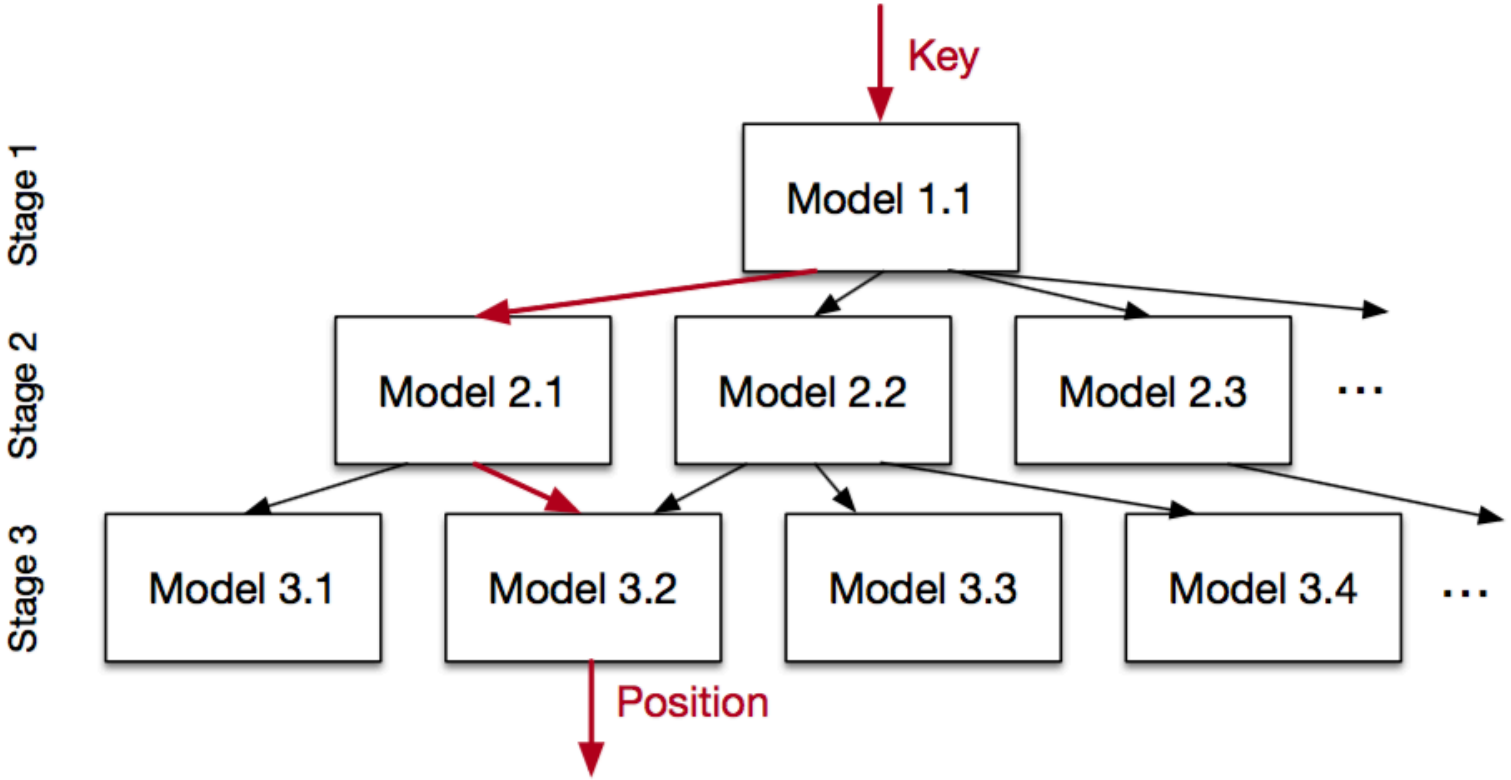


Figure 3: Staged models

https://github.com/yschang1206/dbai-labs

Hands-On Exercises!

- Apply the model to different datasets
- Improve the performance
  - modify the network architecture (e.g., deeper 1-stage model, more 2-stage model), loss function, number of epochs and batch size
  - implement a 3-stage RMI
  - anything you think that would be helpful
- Analyze your results
  - why performance is good/bad
  - what kind of architecture is more suitable for each dataset
  - when keys are not unique or positive?