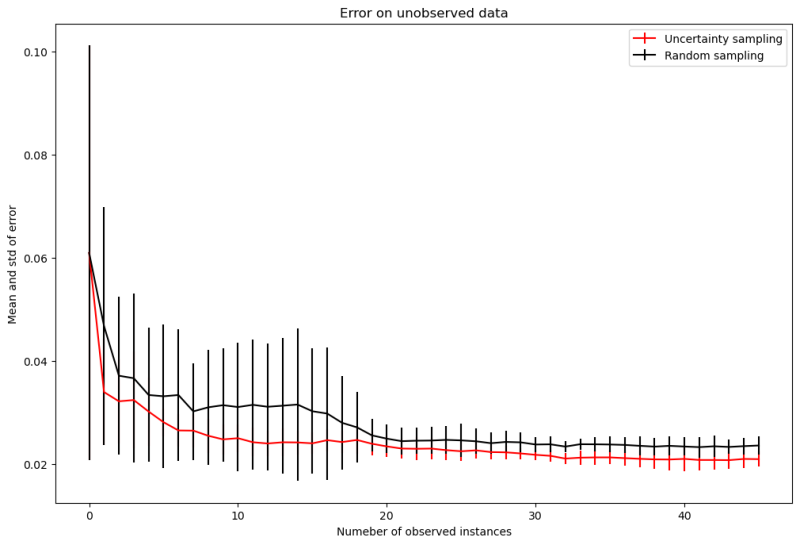
**Simulation:**

1. **Run active learning on dataset from previous research(“previous\_research\_dataset.csv”)**

Start with 5 random initial observed points, query with a batch size = 3 until 50 instances are observed (include initial points, which means query 15 rounds). In each round, record the MSE on unobserved data set. Repeat simulation for 10 times with different initial points. Also plot 10 times random query (without any active learning algorithm). Plot the result by using mean and std of MSE. **Please use seed to make sure that the initial points are the same for active learning query and random query.**

Expected results (a figure copied from 02750 ASR homework):



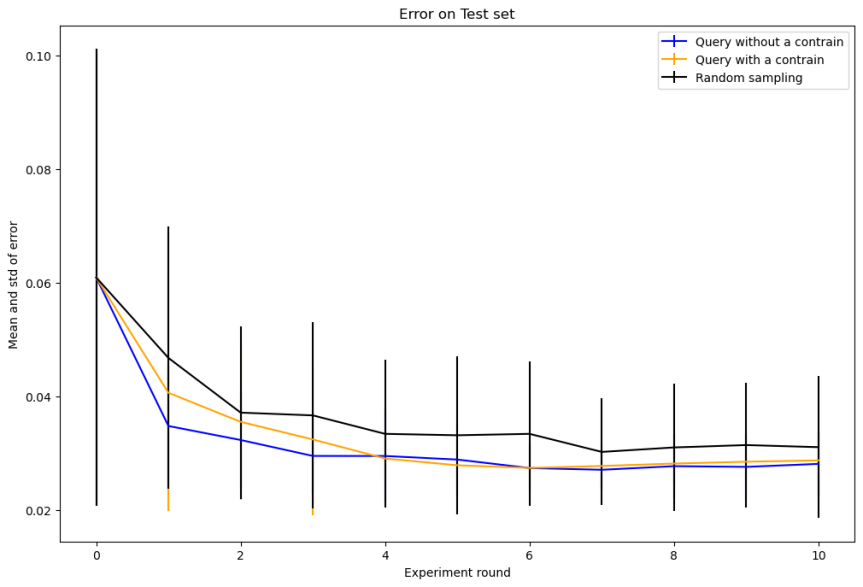
1. **Run active learning on self-generated dataset (“Our\_drug\_simulation\_dataset.csv”)**

**Randomly select 54 data points as the test set.** Start with 54 random initial observed points, query with a batch size = 54 for 10 rounds (if time is too long, just use any number between 5-8). In each round, record the MSE on test set. Repeat simulation for 3 times with different initial points.

Then, query with a constrain that one concentration of one drug will not be queried more than 10 times, 10 query rounds (also, 5-8), 3simulations**.**

Also plot random query (without any active learning algorithm). Plot the result by using mean and std of MSE. **Please use seed to make sure that the initial points are the same for active learning query without a constrain, active learning query with a constrain and random query.**

Expected results:



3. Put 2 figures and a description of your active learning algorithm (what is the base learner for this regression task? What is the query method(uncertainty/density/other)? How to measure uncertainty/density/other metrics for the query method?) into a document.