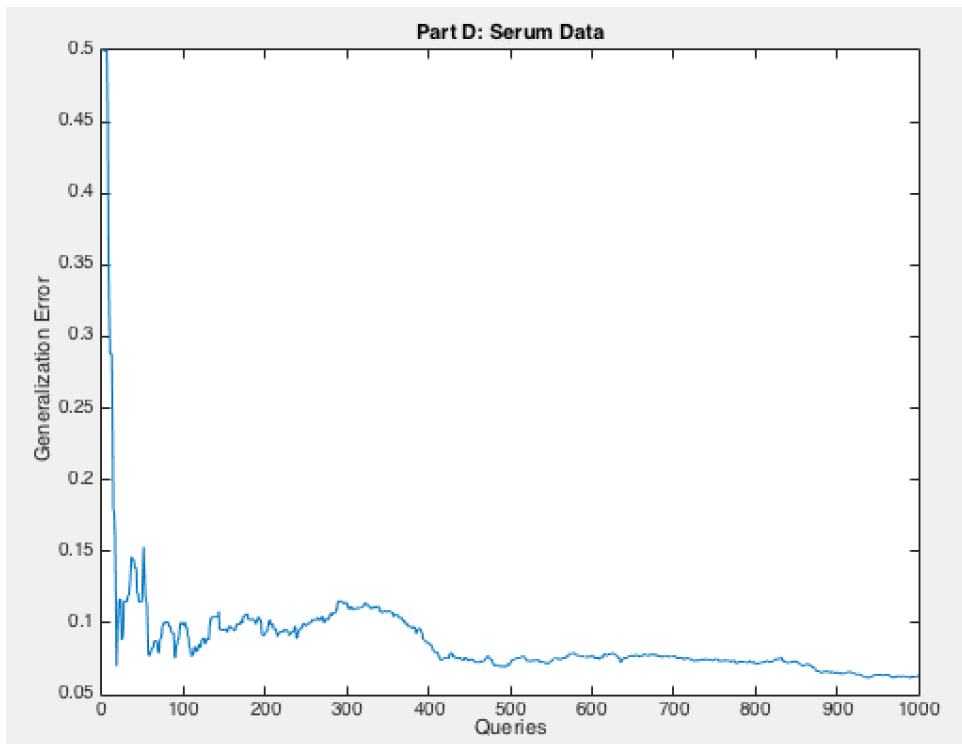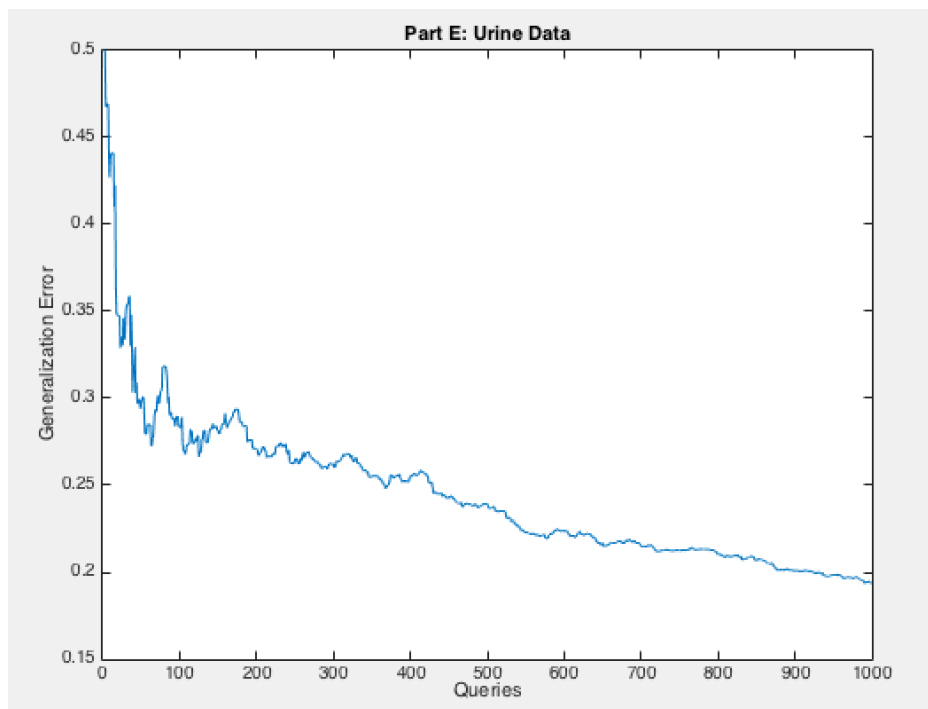**Question1:**

D.

Approximately 400 queries needed before the generalization error converges.
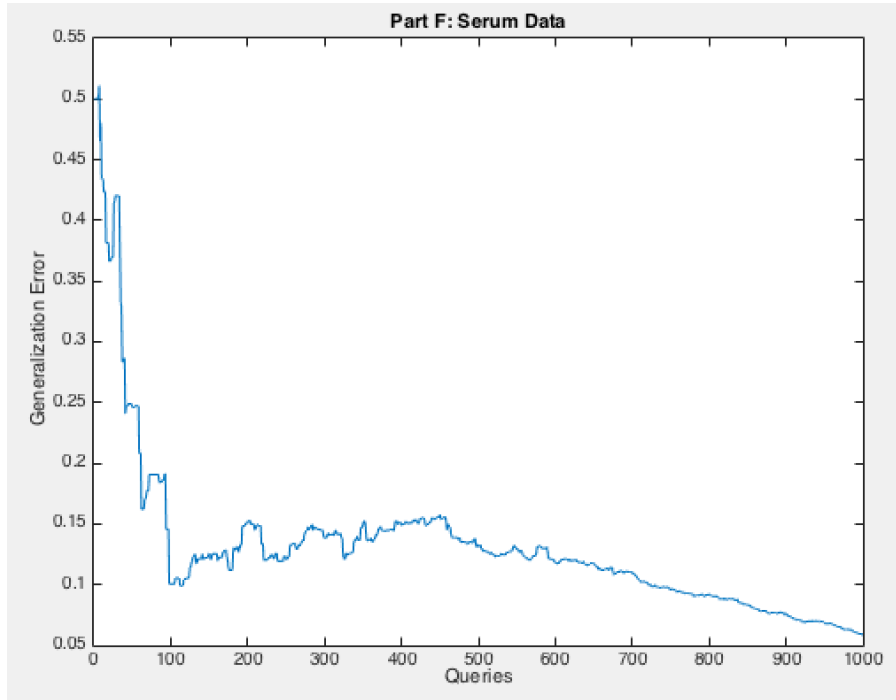
E.

The main differences between D and E are
1. D drops more dramatically at 0-100.
2. The generalization error of D is less than of E.
3. E converges at about 500 queries, which needs more than D.
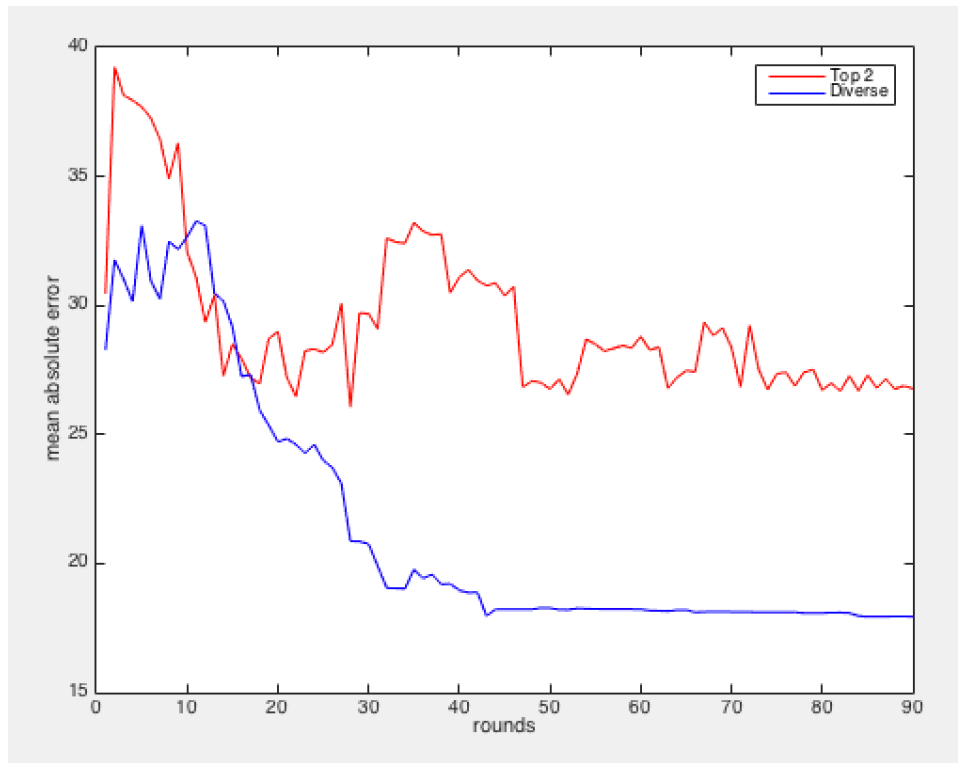
F.

The main differences between D and F are
1. D is more stable after converges.
2. F converges at about 800 queries, which needs more than D.
The strategy for case 2 is more accurate because their results have the similar performance, but D is more stable.

Question2:
B.

I used linear regression to make prediction. And then I calculate the centroid of the clustering of training data, and then use the distance of the point to the centroid as utility score.