

# Tree Based Models

## Decision Trees

- Splits → GINI & entropy
- Uses greedy approach
- Methods to prevent overfitting

Pruning

Bagging

Pros → Explainable Cons → None

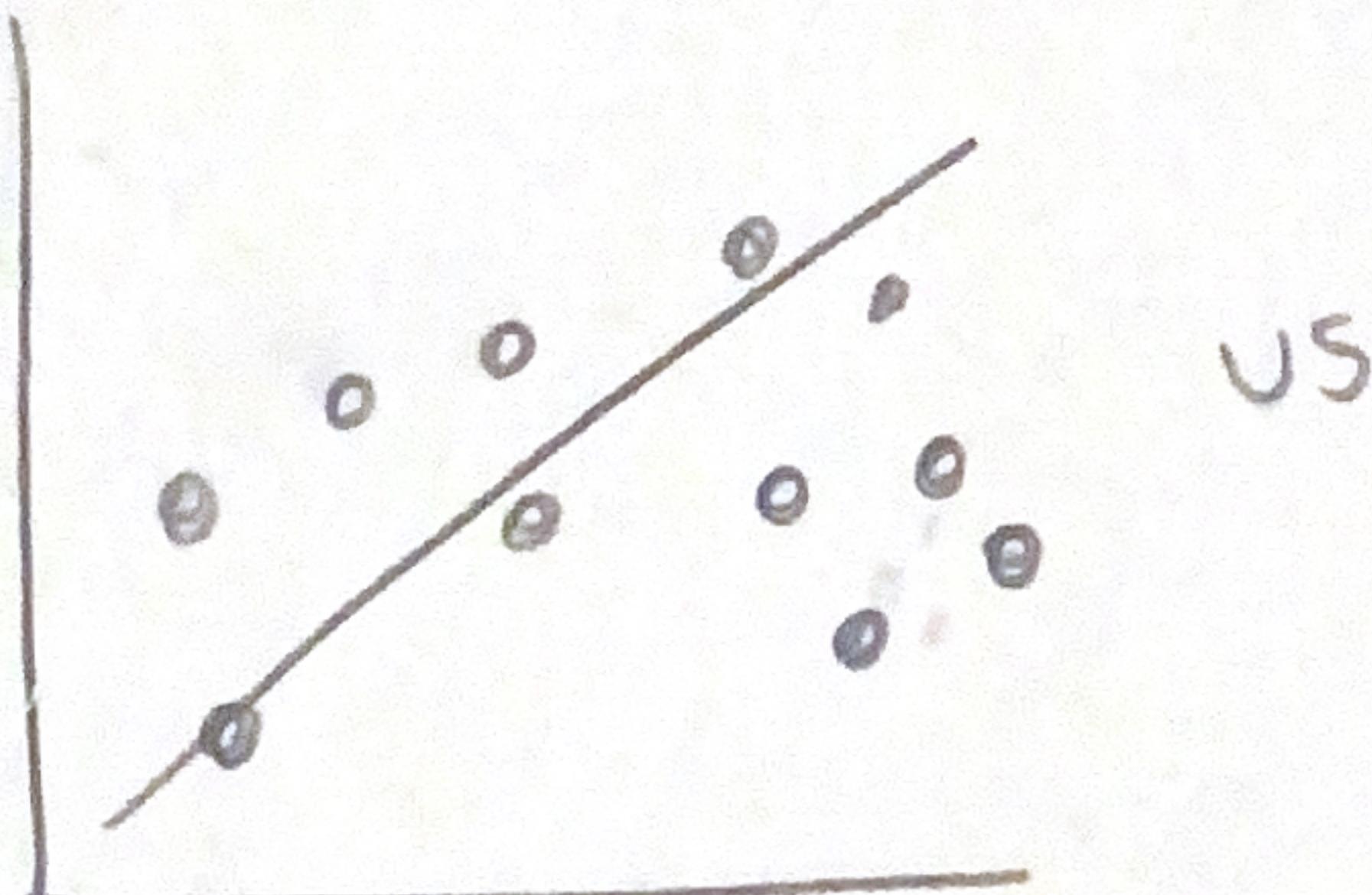
## Ensemble + Trees = Random Forest (Parallel)

- Uses Bagging, decreases variance, random sampling from original data
- Uses randomness: random variables, random bags
- #trees ↑ overfitting ↑

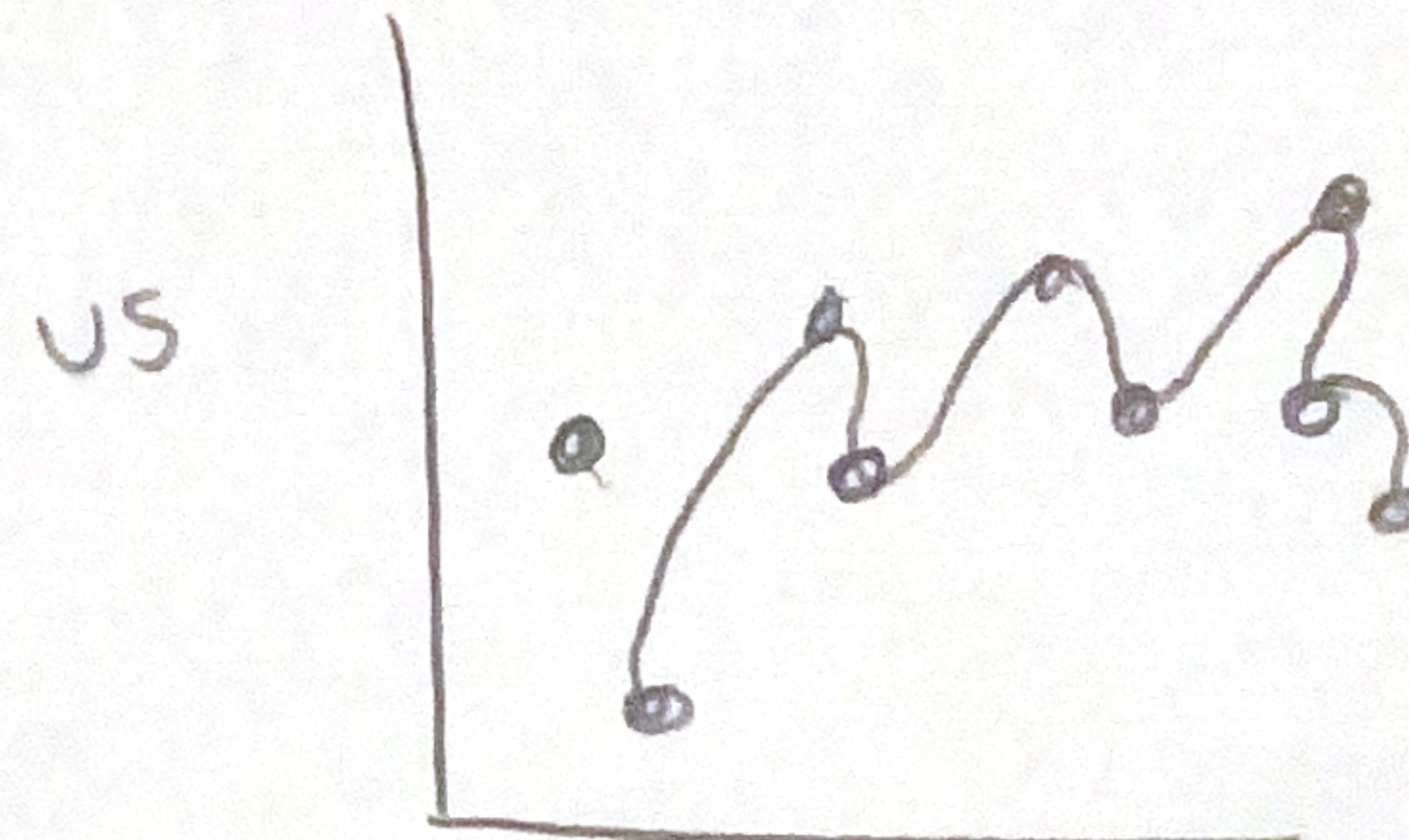
## Serial Ensemble+Trees = Gradient Boosting

Light GBM, AdaBoost, XG Boost, Cat Boost

## Variance-Bias Tradeoff



High Bias  
Underfitting



High Variance  
Overfitting