PGE 383

Ensemble Tree Methods

- Ensemble Methods
- Bootstrap
- Tree Bagging
- Random Forest
- Ensemble Tree Methods Hands-on

PGE 383

Ensemble Tree Methods

- Ensemble Methods

Model Bias and Variance Trade-off

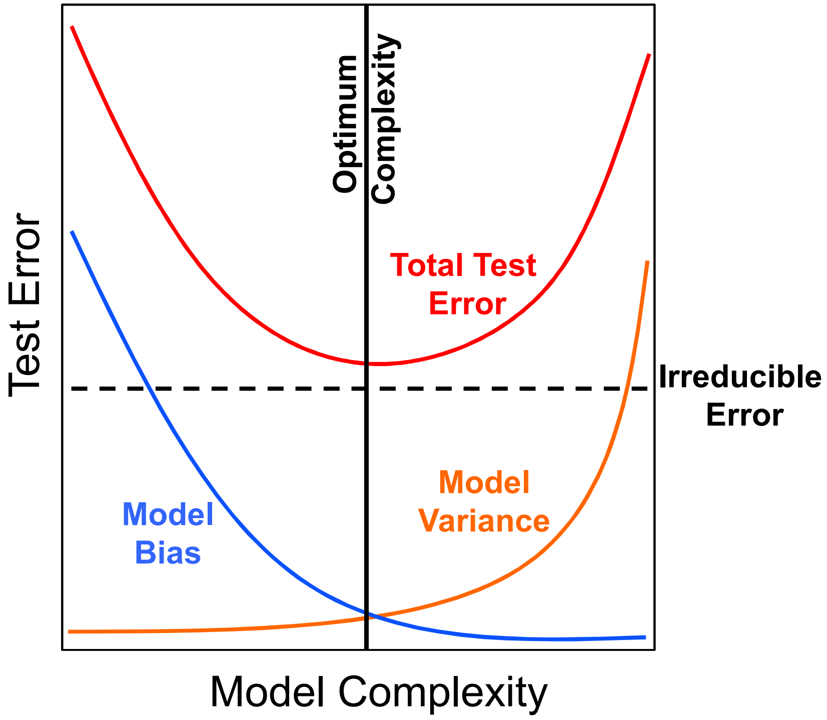
- The **Expected Test Mean Square Error** may be calculated as:

$$\mathrm{E}\left[\left(y_0 - \hat{f}(x_1^0, \dots, x_m^0)\right)^2\right] = \underbrace{Var\left(\hat{f}(x_1^0, \dots, x_m^0)\right)}_{\text{Model Variance}} + \underbrace{\left[Bias\left(\hat{f}(x_1^0, \dots, x_m^0)\right)\right]^2}_{\text{Model Bias}} + \underbrace{Var(\epsilon)}_{\text{Irreducible}}$$

- **Model Variance** is the variance if we had estimated the model with a different training set / sensitivity to data /

- **Model Bias** is error due to using an approximate model / model is too simple

- **Irreducible error** is due to missing variables and limited samples can't be fixed with modeling / entire feature space is not sampled



Reducing Model Variance

- We can improve the **Expected Test Mean Square Error** by reducing the **model variance**:

$$Var\left(\hat{f}\left(x_1^0, \ldots, x_m^0\right)\right)$$

- Averaging is an efficient method to reduce variance:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_s^2}{n}$$

- From standard error we can observe the reduction in variance through averaging.
  - This is maximal when the values averaged are not correlated.

Ensemble Prediction Method

- Calculate multiple estimates for our prediction problem.

- This requires multiple prediction models.

$$\hat{f}^{b}(X_1, \dots, X_m)$$

- We can train multiple models on multiple training datasets.

- But we only have access to a single dataset

$$Y, X_1, \dots, X_m$$

PGE 383
Ensemble Tree Methods

- Bootstrap

Bootstrap Definition

Bootstrap

- method to assess the uncertainty in a sample statistic by repeated random sampling with replacement

Assumptions

- sufficient, representative sampling

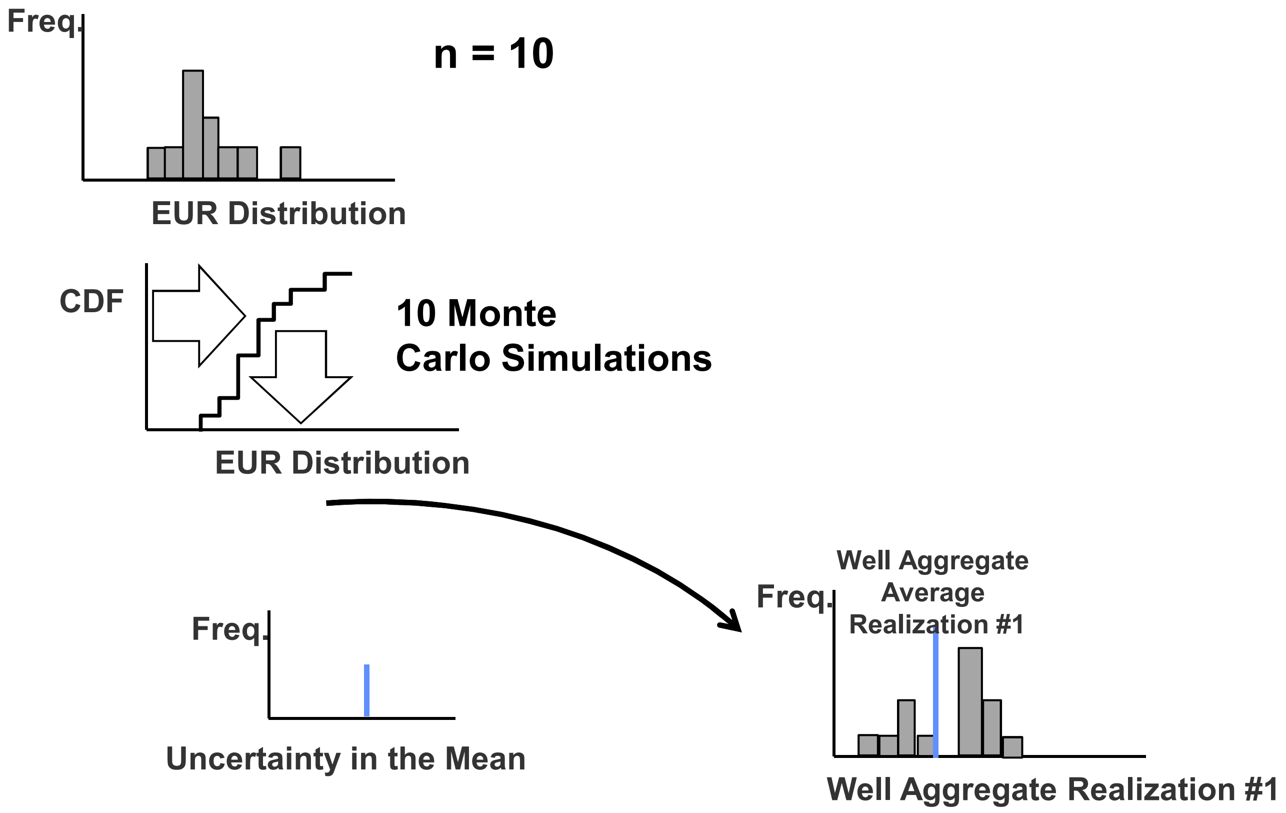
Limitations

- assumes the samples are representative
- assumes stationarity
- only accounts for uncertainty due to too few samples, e.g. no uncertainty due to changes away from data
- does not account for area of interest
- assumes the samples are independent
- does not account for other local information sources

No spatial Context

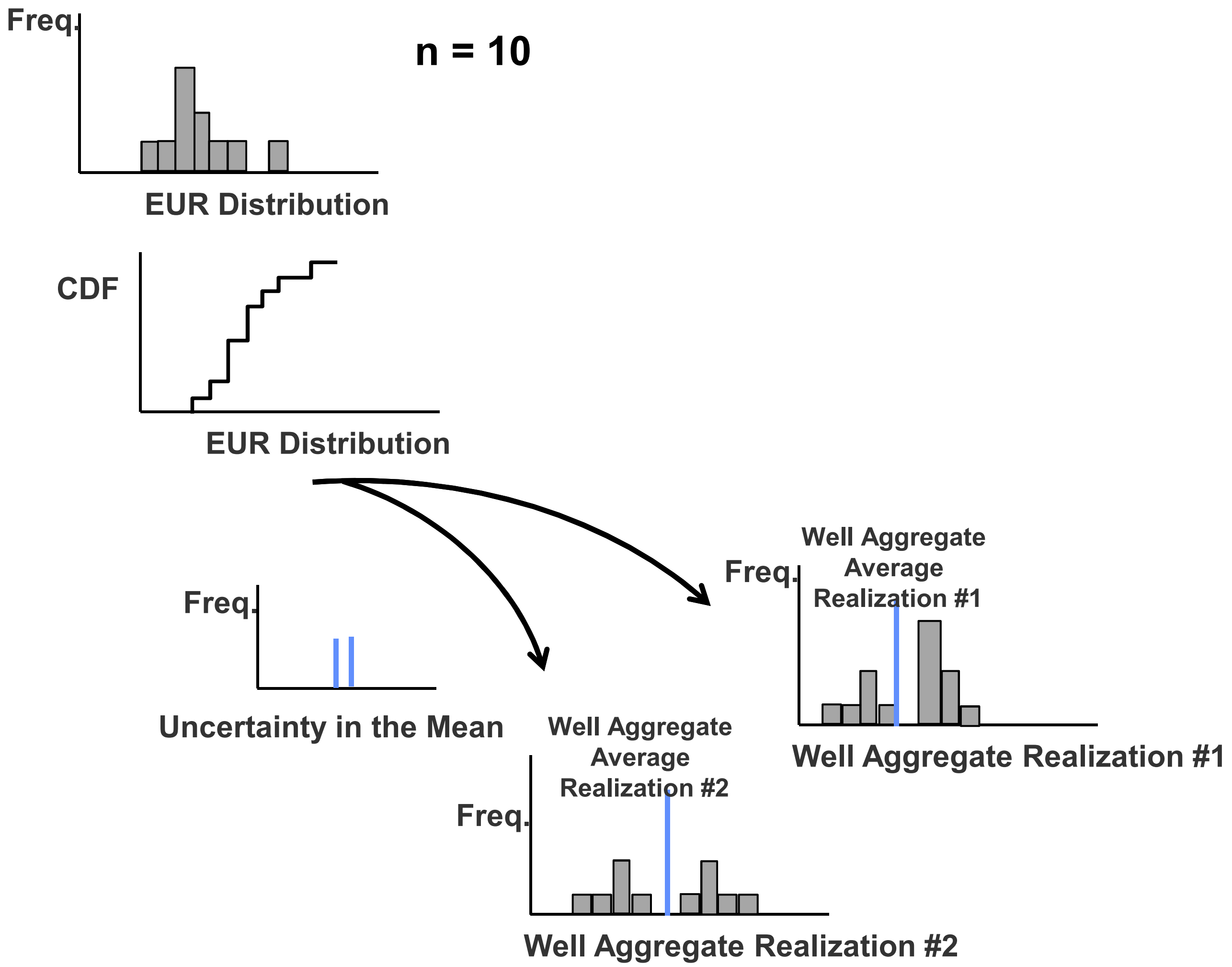
Bootstrap Method

Bootstrap for Uncertainty in the Mean



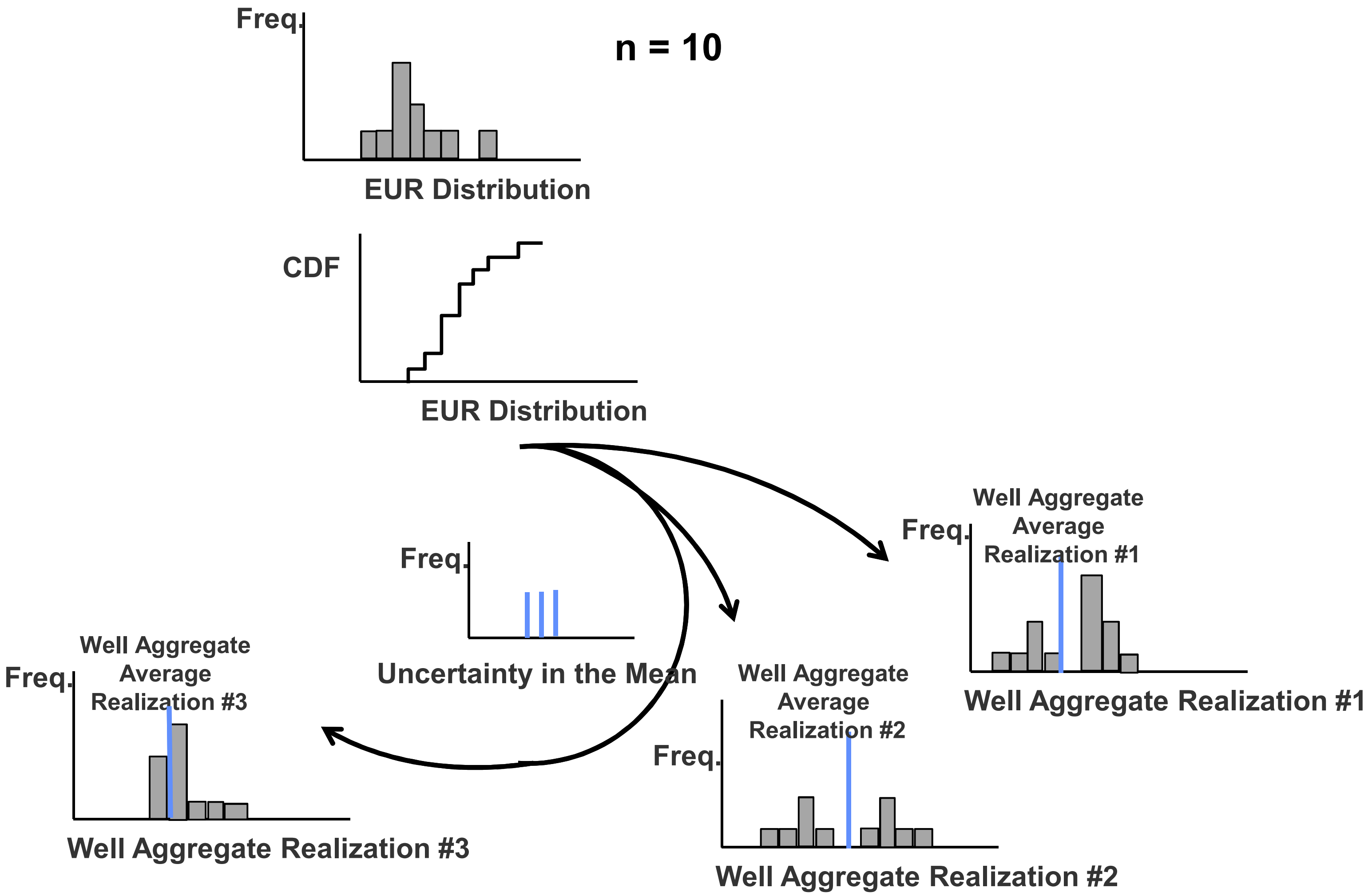
Bootstrap Method

Bootstrap for Uncertainty in the Mean


Freq.
n = 10
EUR Distribution
CDF
EUR Distribution
Freq.
Uncertainty in the Mean
Well Aggregate Average Realization #1
Freq.
Well Aggregate Realization #1
Well Aggregate Average Realization #2
Freq.
Well Aggregate Realization #2

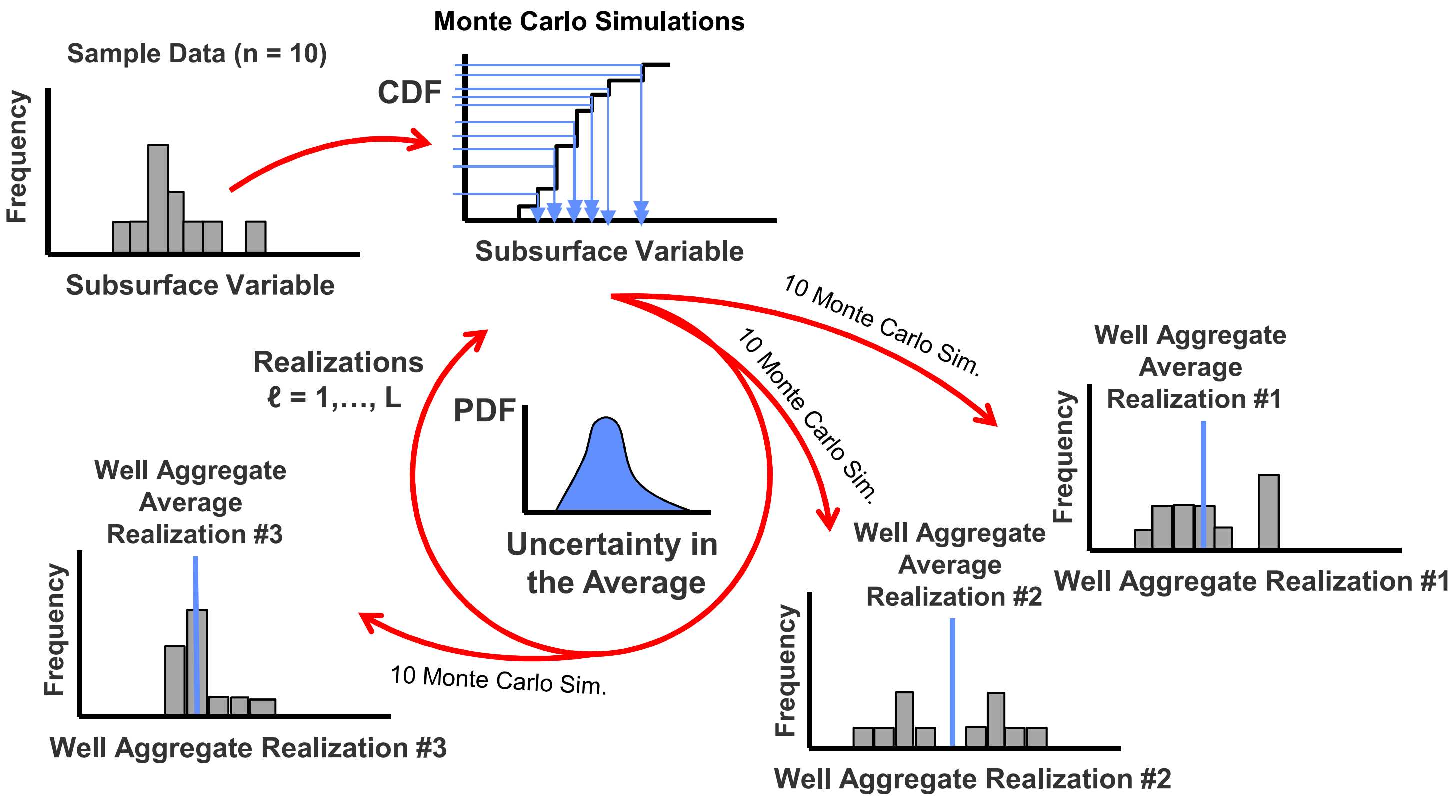
Bootstrap Method

Bootstrap for Uncertainty in the Mean


Freq.
n = 10
EUR Distribution
CDF
EUR Distribution
Freq.
Uncertainty in the Mean
Freq.
Well Aggregate Average Realization #1
Well Aggregate Realization #1
Freq.
Well Aggregate Average Realization #2
Well Aggregate Realization #2
Freq.
Well Aggregate Average Realization #3
Well Aggregate Realization #3

Bootstrap Method

Bootstrap for Uncertainty in the Mean

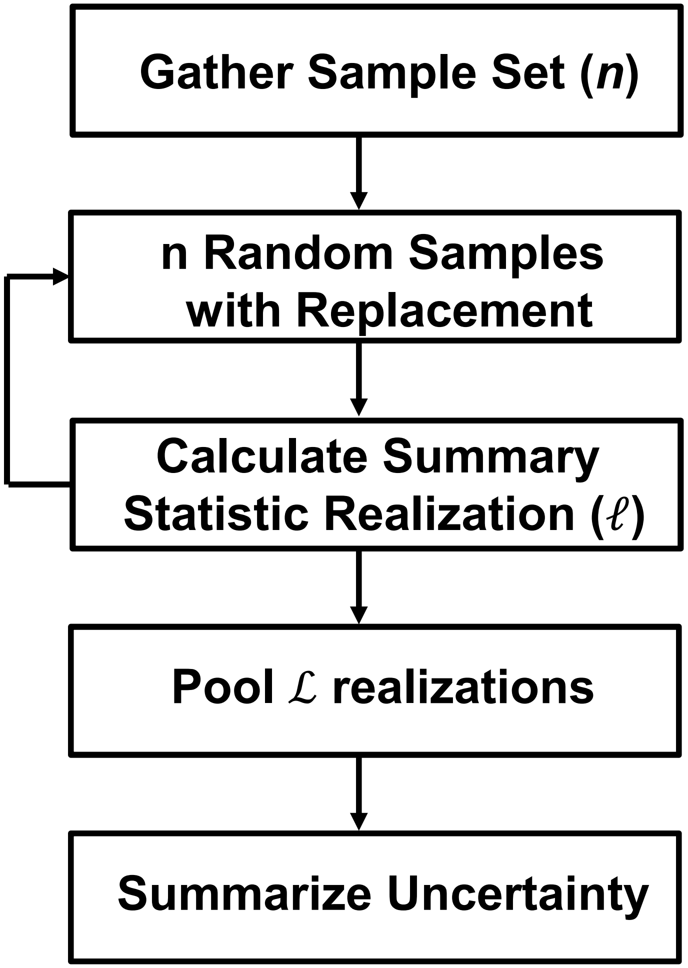

Sample Data (n = 10)
Frequency
Subsurface Variable
Monte Carlo Simulations
CDF
Subsurface Variable
10 Monte Carlo Sim.
10 Monte Carlo Sim.
Well Aggregate Average Realization #1
Frequency
Well Aggregate Realization #1
Realizations
ℓ = 1,…, L
PDF
Uncertainty in the Average
Well Aggregate Average Realization #3
Frequency
Well Aggregate Realization #3
10 Monte Carlo Sim.
Well Aggregate Average Realization #2
Frequency
Well Aggregate Realization #2

Bootstrap Method

- Bootstrap Approach (Efron, 1982)
- Statistical resampling procedure to calculate uncertainty in a calculated statistic from the data itself.
- For uncertainty in the mean solution is standard error:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_s^2}{n}$$

- Extremely powerful. Could get uncertainty in any statistic! e.g. P13, skew etc.
- Would not be possible without bootstrap.
- Advanced forms account for spatial information and strategy (game theory).



Gather Sample Set (*n*)

n Random Samples with Replacement

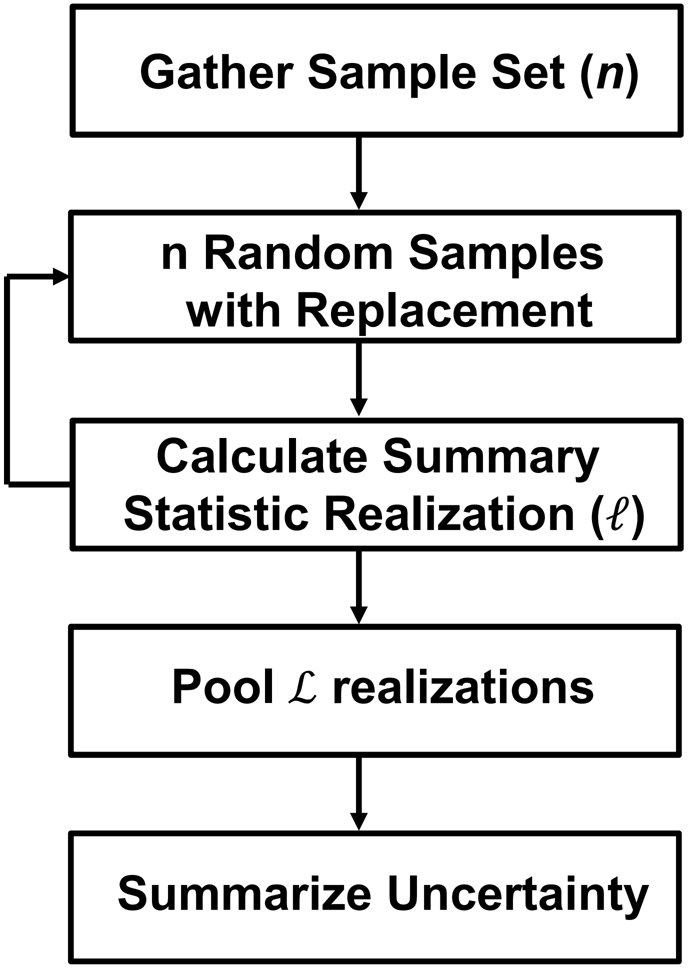
Calculate Summary Statistic Realization (ℓ)

Pool $\mathcal{L}$ realizations

Summarize Uncertainty

Bootstrap Method

- You now know about one of the most powerful tools ever!
- Caveats:
  - assumes the sample set is representative
  - unbiased and covers the full range
  - assumes all samples are independent if not consider Journel’s spatial bootstrap (1993).
- You can do bootstrap in Excel.



Gather Sample Set (n)

n Random Samples with Replacement

Calculate Summary Statistic Realization (ℓ)

Pool $\mathcal{L}$ realizations

Summarize Uncertainty

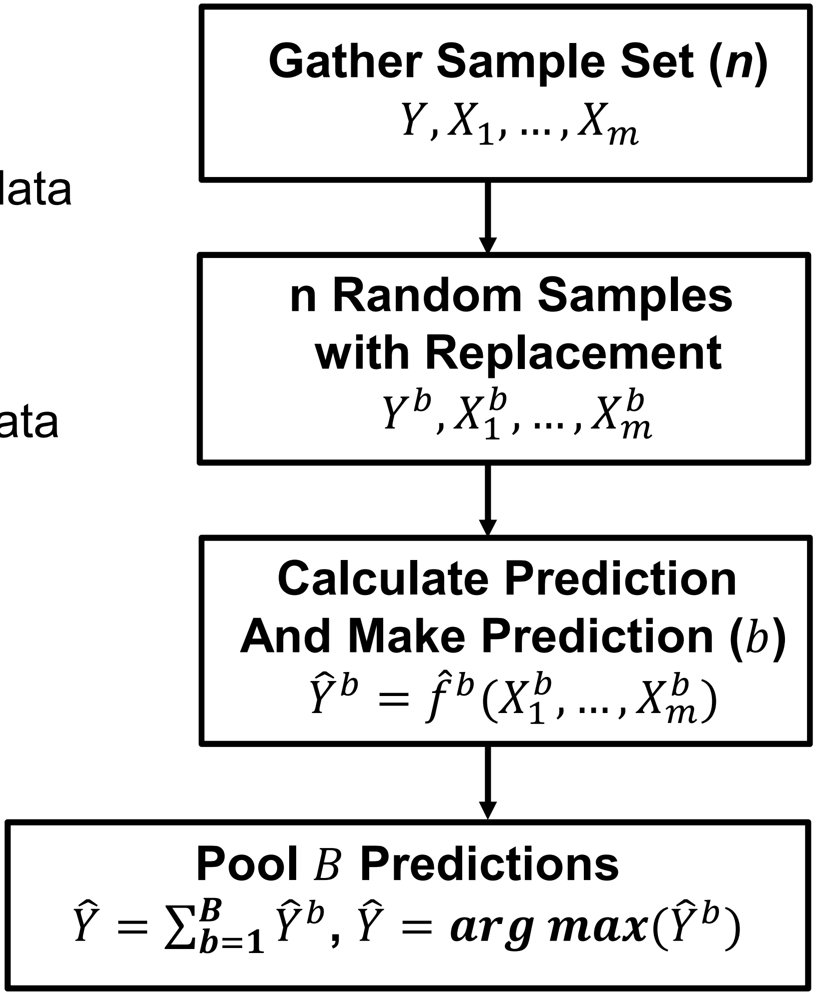
Bootstrap
For Ensemble Methods

1. Apply statistical bootstrap to obtain multiple realizations of the training data

$$Y^b, X_1^b, \dots, X_m^b, b = 1, \dots, B$$

2. Build a prediction model for each data realization

$$\hat{Y}^b = \hat{f}^b(X_1^b, \dots, X_m^b)$$



Gather Sample Set (*n*)
$Y, X_1, \dots, X_m$

n Random Samples with Replacement
$Y^b, X_1^b, \dots, X_m^b$

Calculate Prediction And Make Prediction (b)
$\hat{Y}^b = \hat{f}^b(X_1^b, \dots, X_m^b)$

Pool B Predictions
$\hat{Y} = \sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{Y}^b, \hat{Y} = \boldsymbol{arg\,max}(\hat{Y}^b)$

Bootstrap
For Ensemble Methods

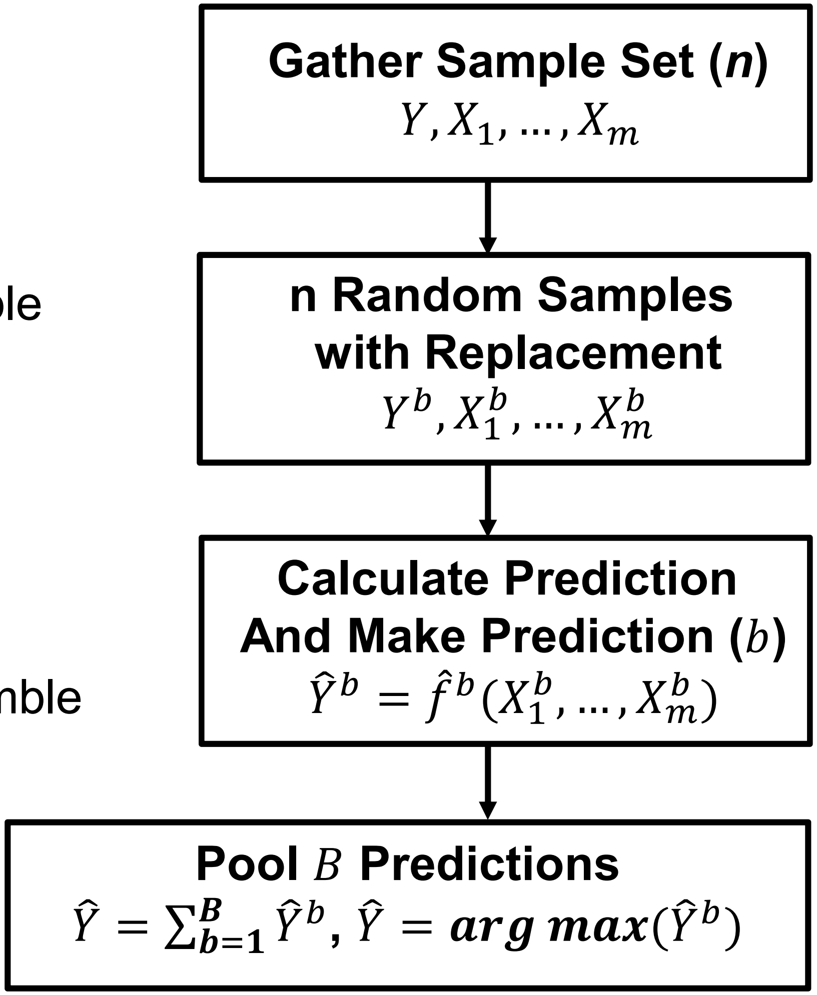
3. Apply all prediction models for each prediction:

- Regression – aggregate the ensemble predictions with the average

$$\hat{Y} = \sum_{b=1}^{B} \hat{Y}^{b}$$

- Classification – aggregate the ensemble predictions with majority-rule

$$\hat{Y} = \boldsymbol{arg\,max}(\hat{Y}^{b})$$



Gather Sample Set (*n*)

$Y, X_1, \ldots, X_m$

n Random Samples with Replacement

$Y^b, X_1^b, \ldots, X_m^b$

Calculate Prediction And Make Prediction (*b*)

$\hat{Y}^b = \hat{f}^b(X_1^b, \ldots, X_m^b)$

Pool *B* Predictions

$\hat{Y} = \sum_{\boldsymbol{b=1}}^{\boldsymbol{B}} \hat{Y}^b, \hat{Y} = \boldsymbol{arg\,max}(\hat{Y}^b)$

PGE 383 Lecture 26

Ensemble Tree Methods

- **Tree Bagging**

Tree Bagging

Build an ensemble of decision trees with multiple, bootstrap realizations of the data.

Comments:

- The ensemble approach **reduces model variance as expected**

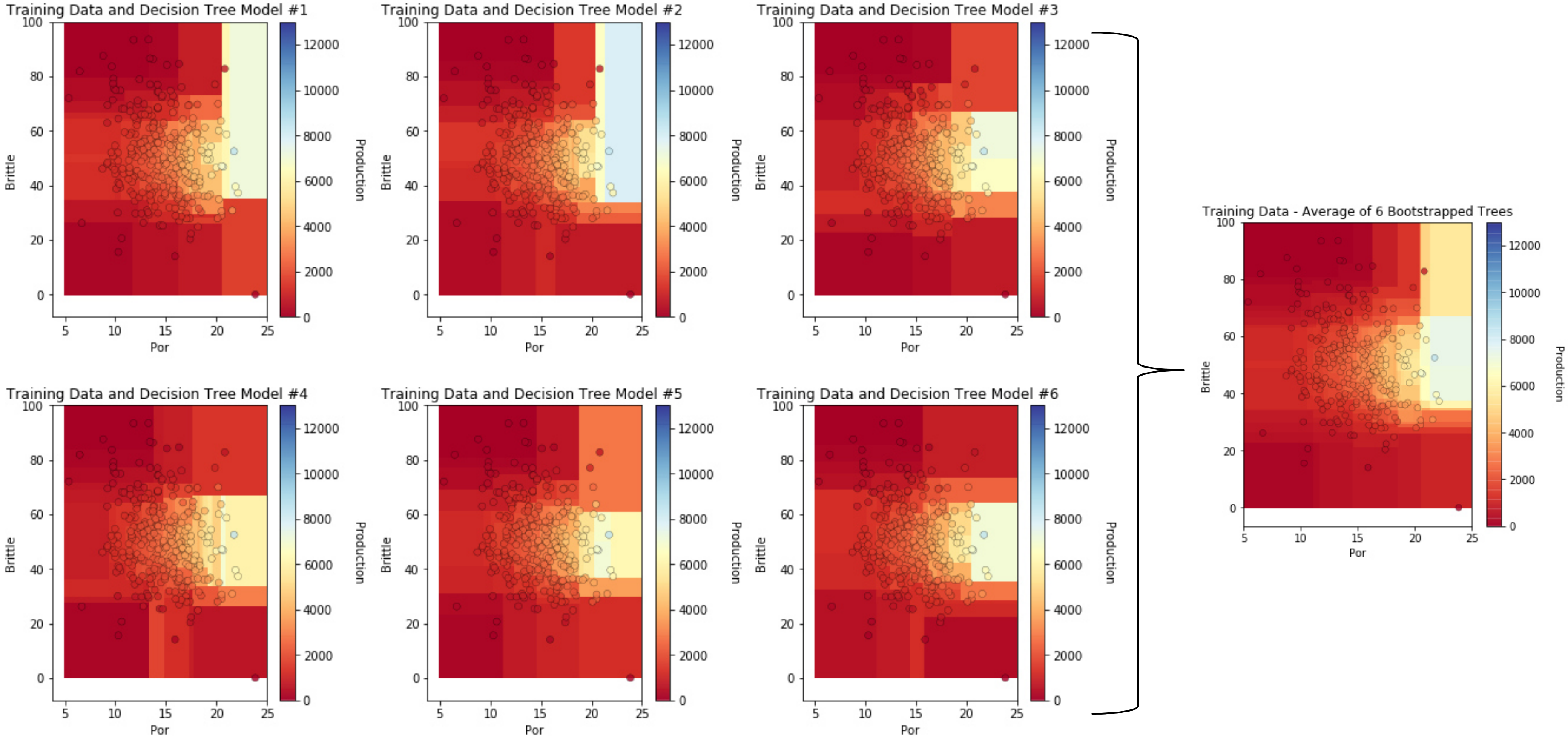
- **Grow each tree to be complicated.** We can overfit each tree, the ensemble approach reduces the risk of overfit

- In expectation, 1/3 of the data is not used for each tree, this provides the opportunity to have access to out-of-bag samples for cross validation, so **we can build our model and cross validate with all the data at once**

- We want the trees to be decorrelated, diverse to maximize the reduction in model variance, **this leads to random forest**

Tree Bagging Example

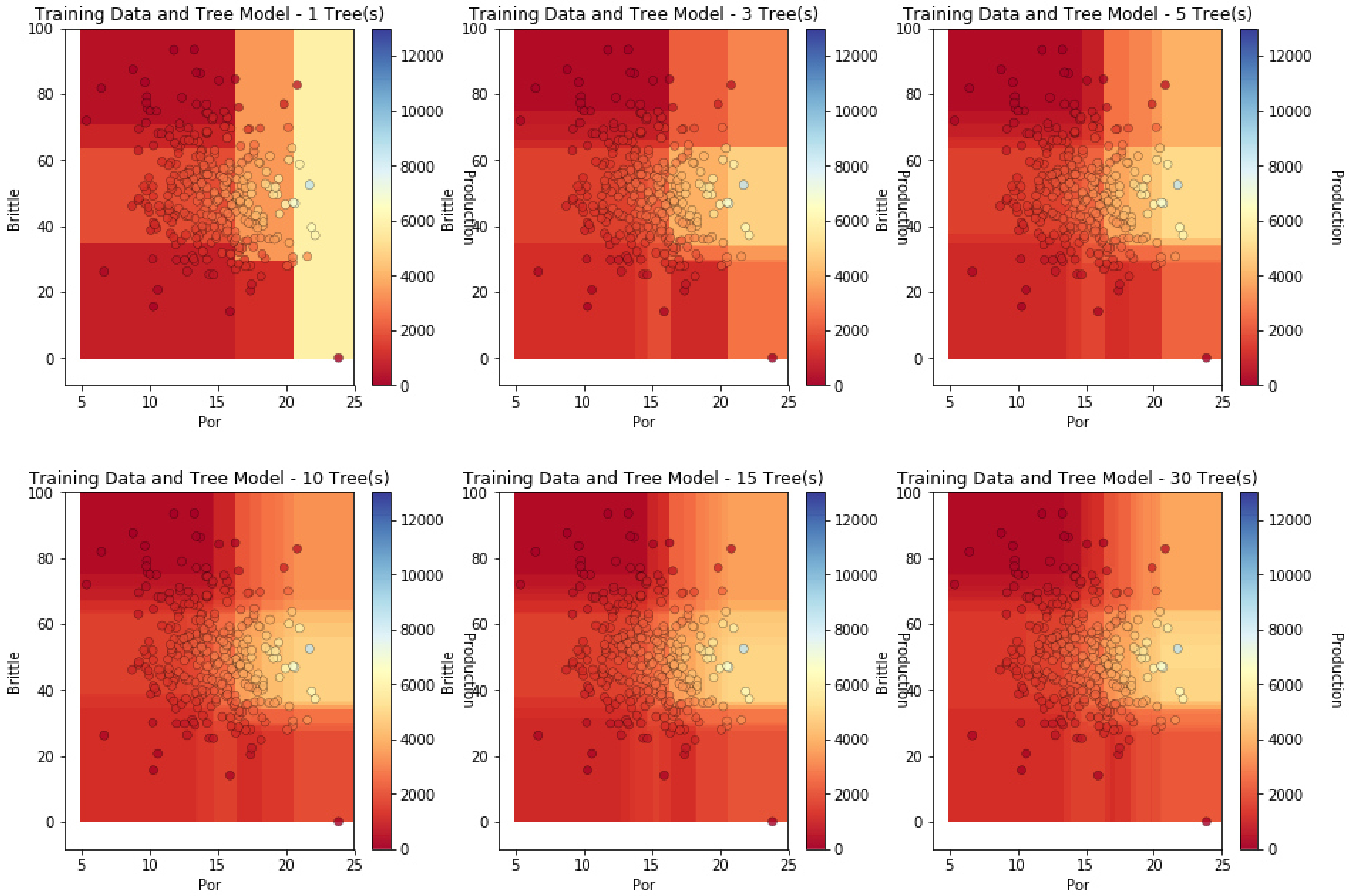
Build an ensemble of decision trees with multiple, bootstrap realizations of the data and average the predictions from all models.



6 bootstrapped, complicated decision trees (left) and the average of all 6 models (right).

Tree Bagging Example

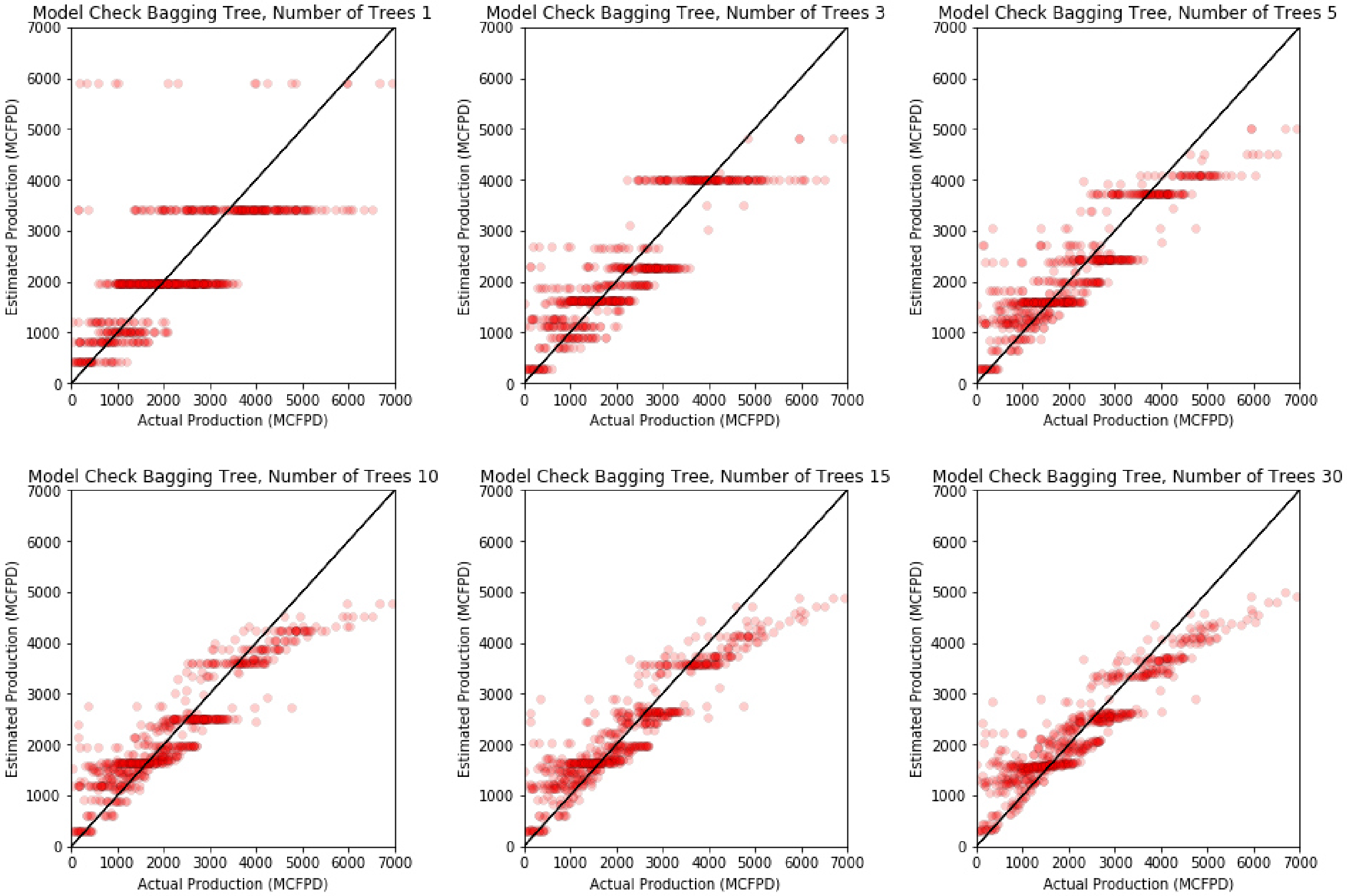
Observe the impact on the prediction model with the addition of more trees – transition from a discontinuous to continuous prediction model!



6 tree bagging prediction models and all training data with increasing number of trees.

Tree Bagging Example

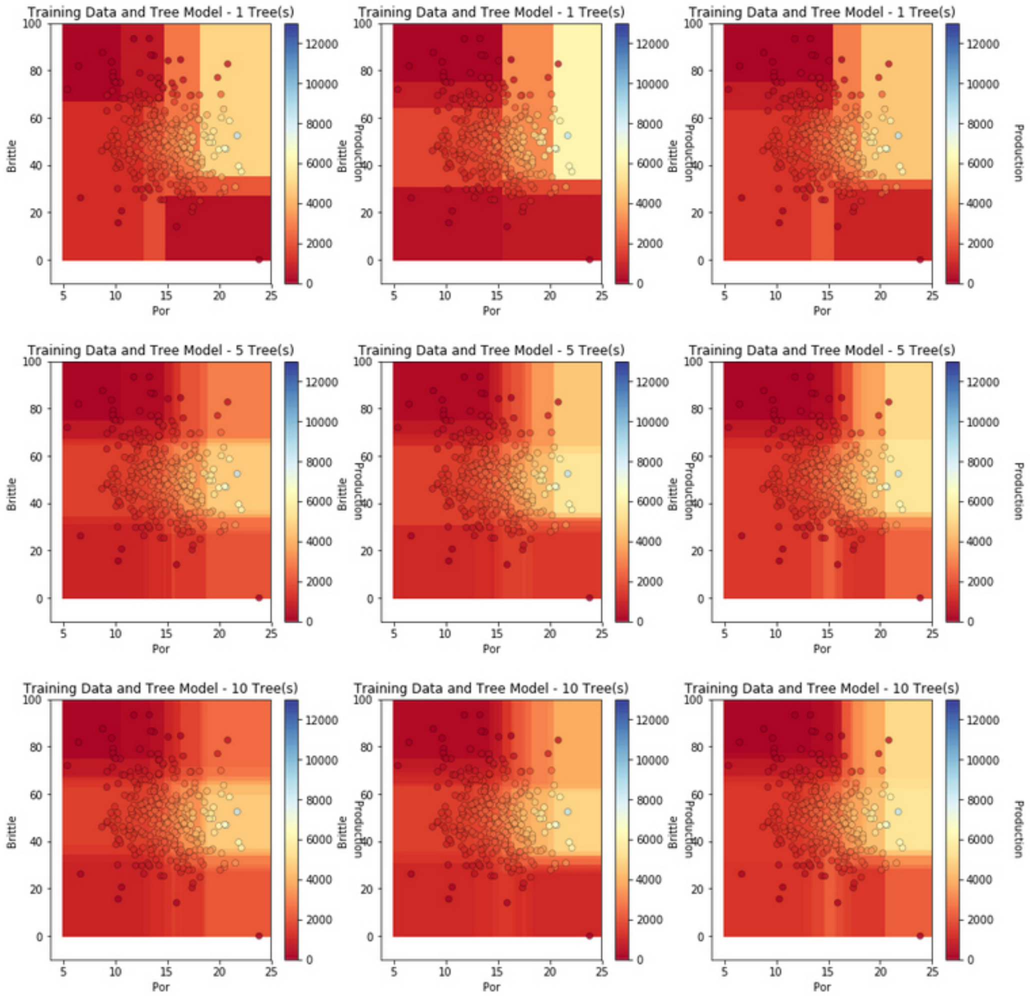
Observe the improved testing accuracy in cross validation with increasing number of trees.



Cross validation with 6 tree bagging prediction models with increasing number of trees.

Tree Bagging Example

Observe the reduction in model variance with increasing number of trees.



3 models with 1, 5 and 10 trees to demonstrate the reduction in model variance with increased ensemble aggregation.

PGE 383

Ensemble Tree Methods

- Random Forest

Random Forest

A limitation with tree bagging is that the individual trees may be highly correlated

- This occurs when there is a dominant predictor feature as it will always be applied to the top split(s)
  - the result is all the trees in the ensemble are very similar (i.e. correlated)

- With highly correlated trees, there is significantly less reduction in model variance with the ensemble

- Random forest is tree bagging and for each split only a subset of the m available predictors are candidates for splits (selected at random)

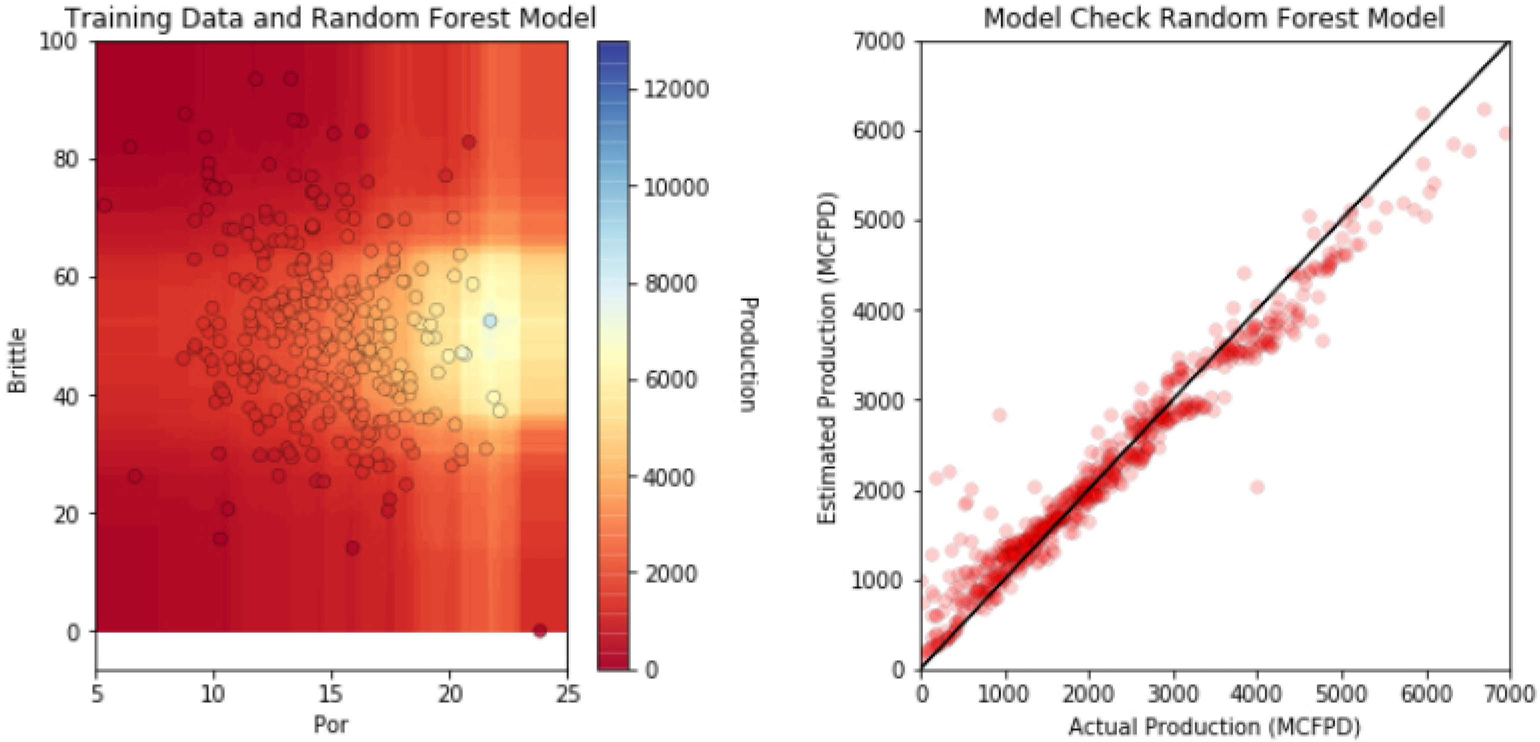
$$p = \sqrt{m}$$

- This forces each tree in the ensemble to evolve in dissimilar manner

Random Forest Example

Example random forest model for the previous prediction problem

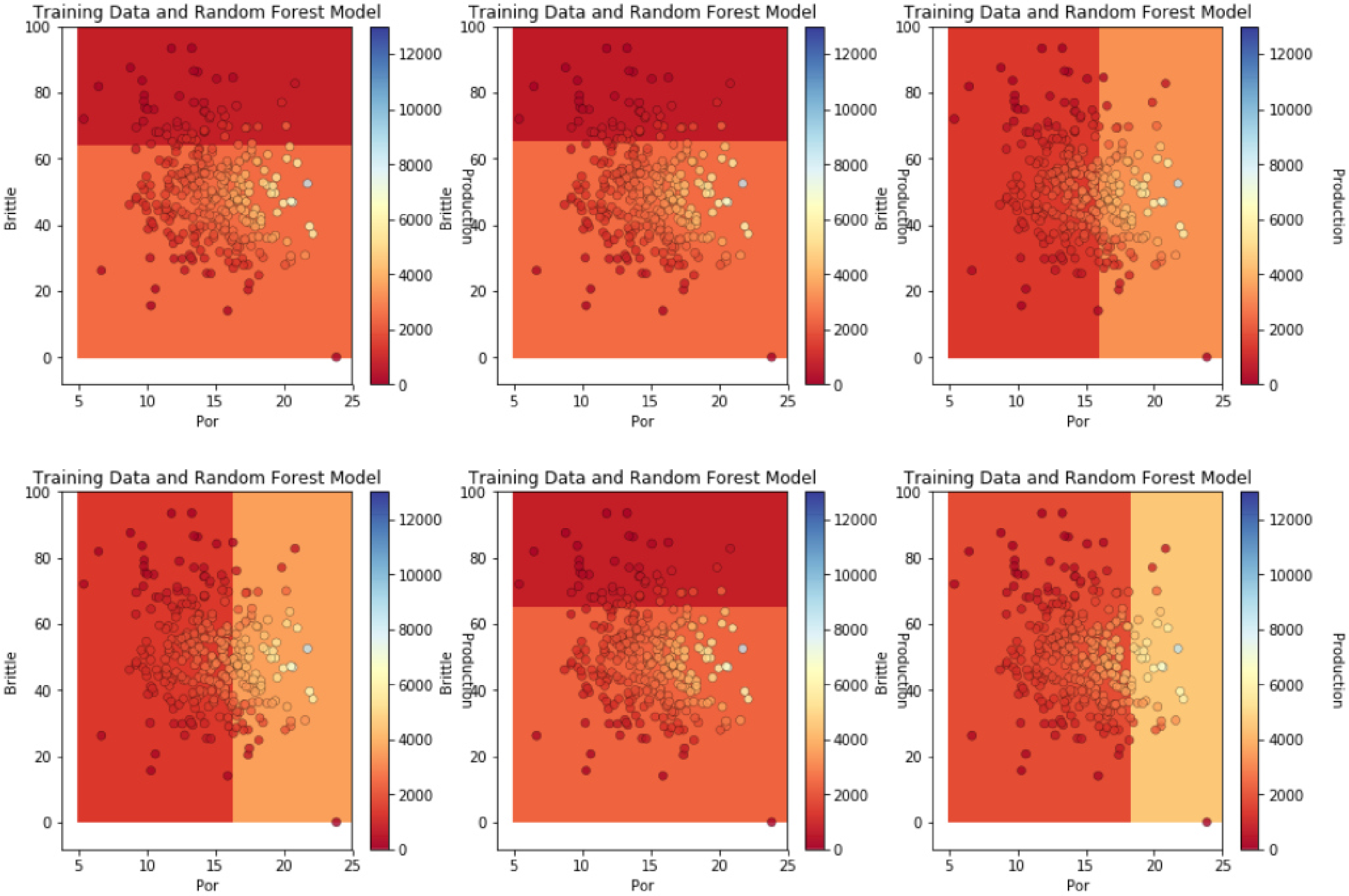
- 300 trees, trained to a maximum depth of 3, 1 predictor selected for each split



Random forest prediction model and model cross validation.

Random Forest Example

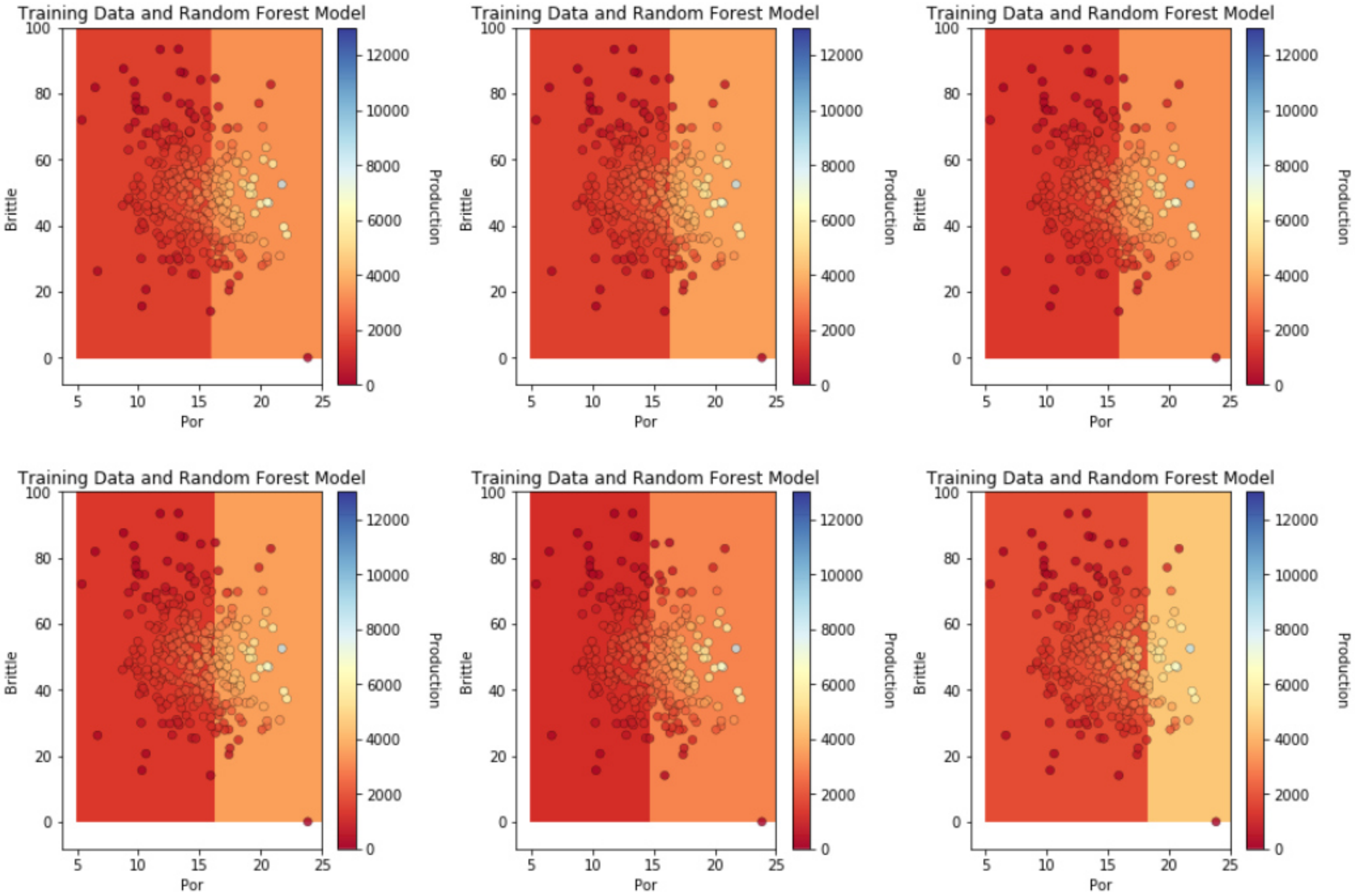
Are the trees diverse? Let’s freeze 6 trees at the first split.



6 tree from a random forest frozen at the first split (p = 1)

Random Forest Example

Now compare to tree bagging, just set $p = m$ to get tree bagging from random forest! There is very little tree diversity on the first split!


Training Data and Random Forest Model

Training Data and Random Forest Model

Training Data and Random Forest Model

Training Data and Random Forest Model

Training Data and Random Forest Model

Training Data and Random Forest Model

6 tree from tree bagging frozen at the first split (p = 1)

Out-of-Bag Cross Validation Hyperparameter Tuning

During the construction of B decision trees (tree bagging and random forest)

- Build a decision tree with the 2/3 of data (in expectation) sampled with bootstrap
  - The data not used in the current model training are know as out-of-bag samples

  - Predict at the out-of-bag samples

- Pool the B/3 predictions for each sample data from all the B models and make an out-of-bag prediction

$$\hat{\boldsymbol{y}}_{\boldsymbol{\alpha}}^{*} = \sum_{\boldsymbol{b}=\boldsymbol{1}}^{\frac{\boldsymbol{B}}{\boldsymbol{3}}} \hat{y}_{\boldsymbol{\alpha}}^{*,\boldsymbol{b}}$$

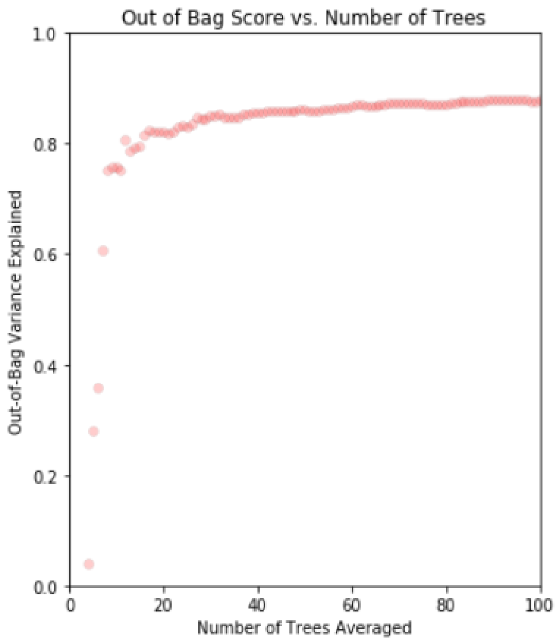
- Calculate the out-of-bag mean square error to access model performance.

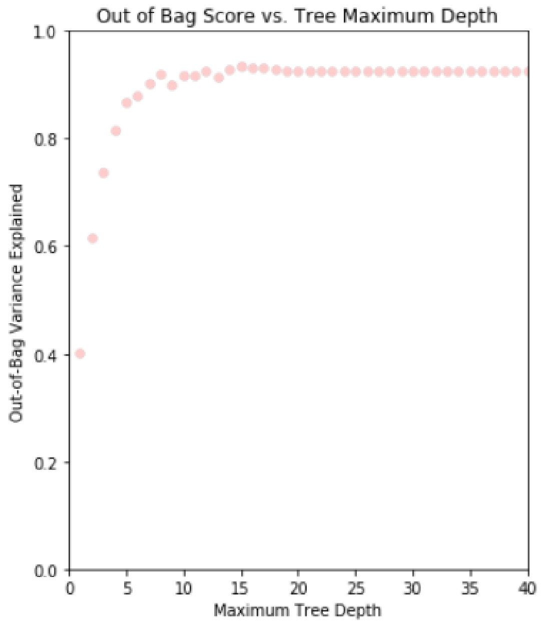
$$MSE_{OOB} = \sum_{\boldsymbol{b}=\boldsymbol{1}}^{\frac{\boldsymbol{B}}{\boldsymbol{3}}} [\hat{\boldsymbol{y}}_{\boldsymbol{\alpha}}^{*} - \boldsymbol{y}_{\boldsymbol{\alpha}}]^{\boldsymbol{2}}$$

Out-of-Bag Cross Validation Hyperparameter Tuning

For previous random forest

- Loop over multiple hyperparameters and calculate the out-of-bag prediction performance





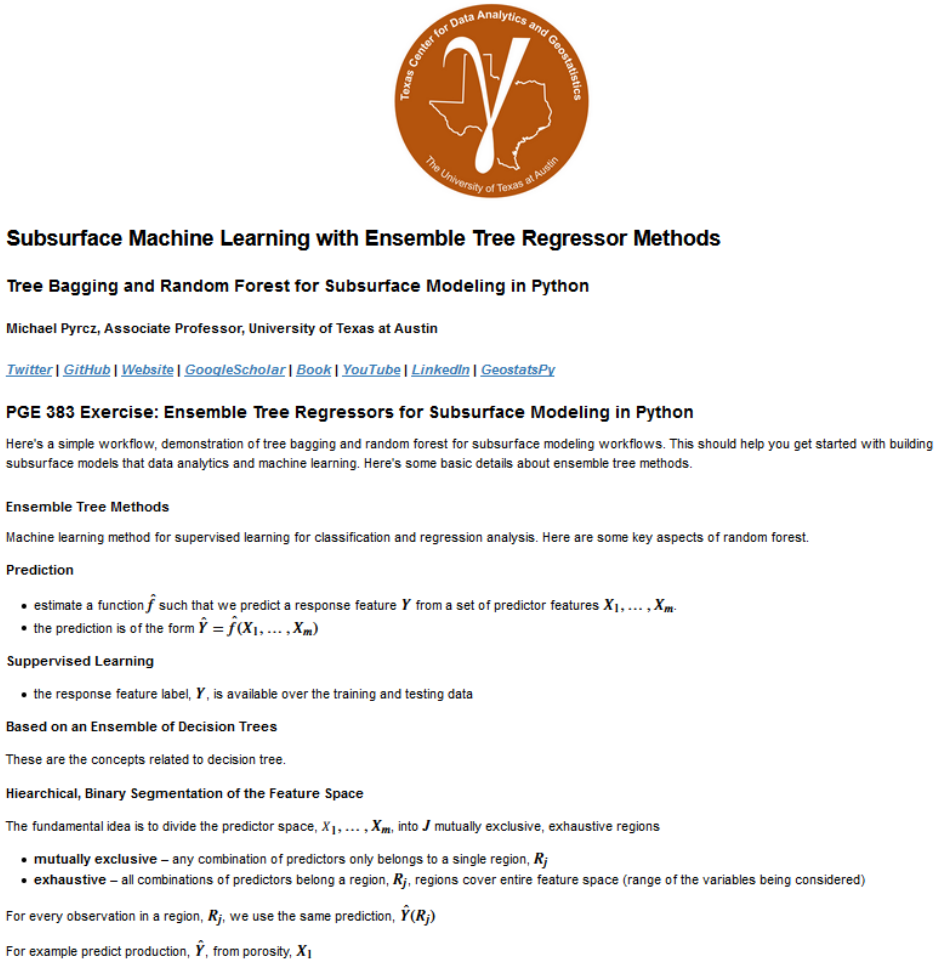
- note that our random forest model is robust and resistant to overfit, the out-of-bag performance evaluation is approximately monotonically increasing

PGE 383
Ensemble Tree Methods

- **Ensemble Tree Methods Hands-on**

Ensemble Tree Demonstration

Demonstration workflow with ensemble tree regression.



Subsurface Machine Learning with Ensemble Tree Regressor Methods

Tree Bagging and Random Forest for Subsurface Modeling in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

Twitter | GitHub | Website | GoogleScholar | Book | YouTube | LinkedIn | GeostatsPy

PGE 383 Exercise: Ensemble Tree Regressors for Subsurface Modeling in Python

Here's a simple workflow, demonstration of tree bagging and random forest for subsurface modeling workflows. This should help you get started with building subsurface models that data analytics and machine learning. Here's some basic details about ensemble tree methods.

Ensemble Tree Methods

Machine learning method for supervised learning for classification and regression analysis. Here are some key aspects of random forest.

Prediction

- estimate a function $\hat{f}$ such that we predict a response feature Y from a set of predictor features $X_1, \ldots, X_m$.
- the prediction is of the form $\hat{Y} = \hat{f}(X_1, \ldots, X_m)$

Suppervised Learning

- the response feature label, Y, is available over the training and testing data

Based on an Ensemble of Decision Trees

These are the concepts related to decision tree.

Hiearchical, Binary Segmentation of the Feature Space

The fundamental idea is to divide the predictor space, $X_1, \ldots, X_m$, into J mutually exclusive, exhaustive regions

- **mutually exclusive** – any combination of predictors only belongs to a single region, R_j
- **exhaustive** – all combinations of predictors belong a region, R_j, regions cover entire feature space (range of the variables being considered)

For every observation in a region, R_j, we use the same prediction, $\hat{Y}(R_j)$

For example predict production, $\hat{Y}$, from porosity, X_1

File SubsurfaceDataAnalytics_EnsembleTree.ipynb at https://git.io/fjX23.

PGE 383

Ensemble Tree Methods

- Ensemble Methods
- Bootstrap
- Tree Bagging
- Random Forest
- Ensemble Tree Methods Hands-on