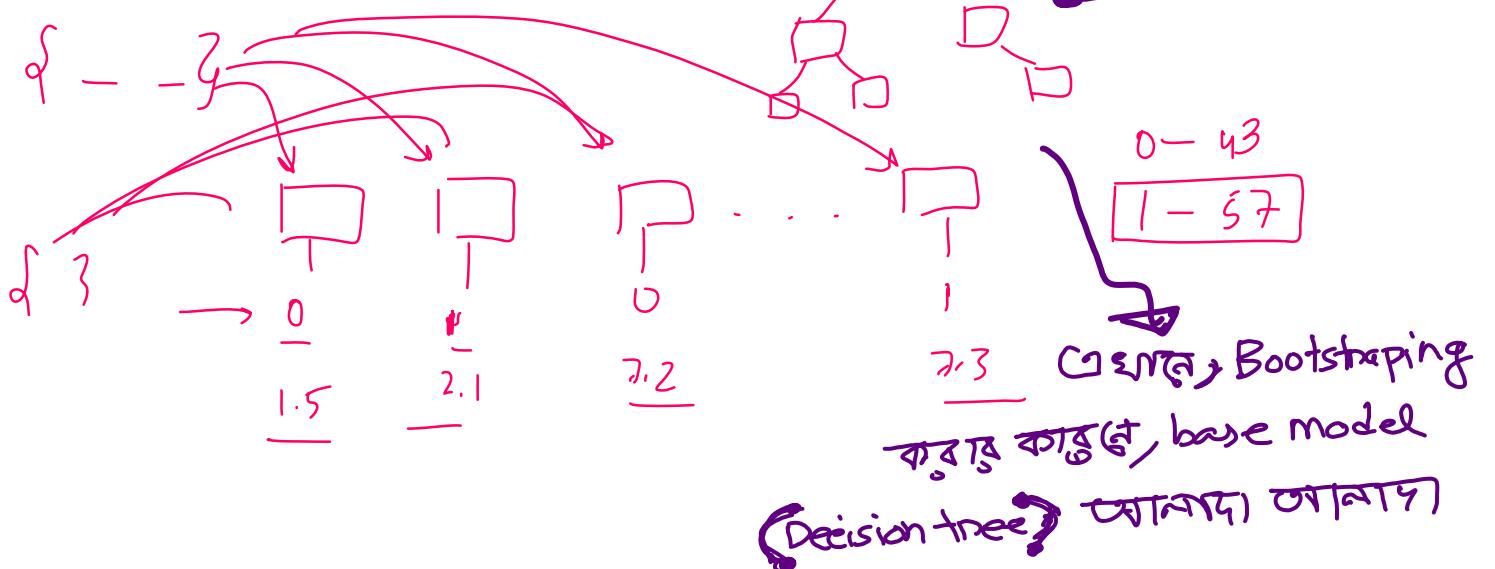
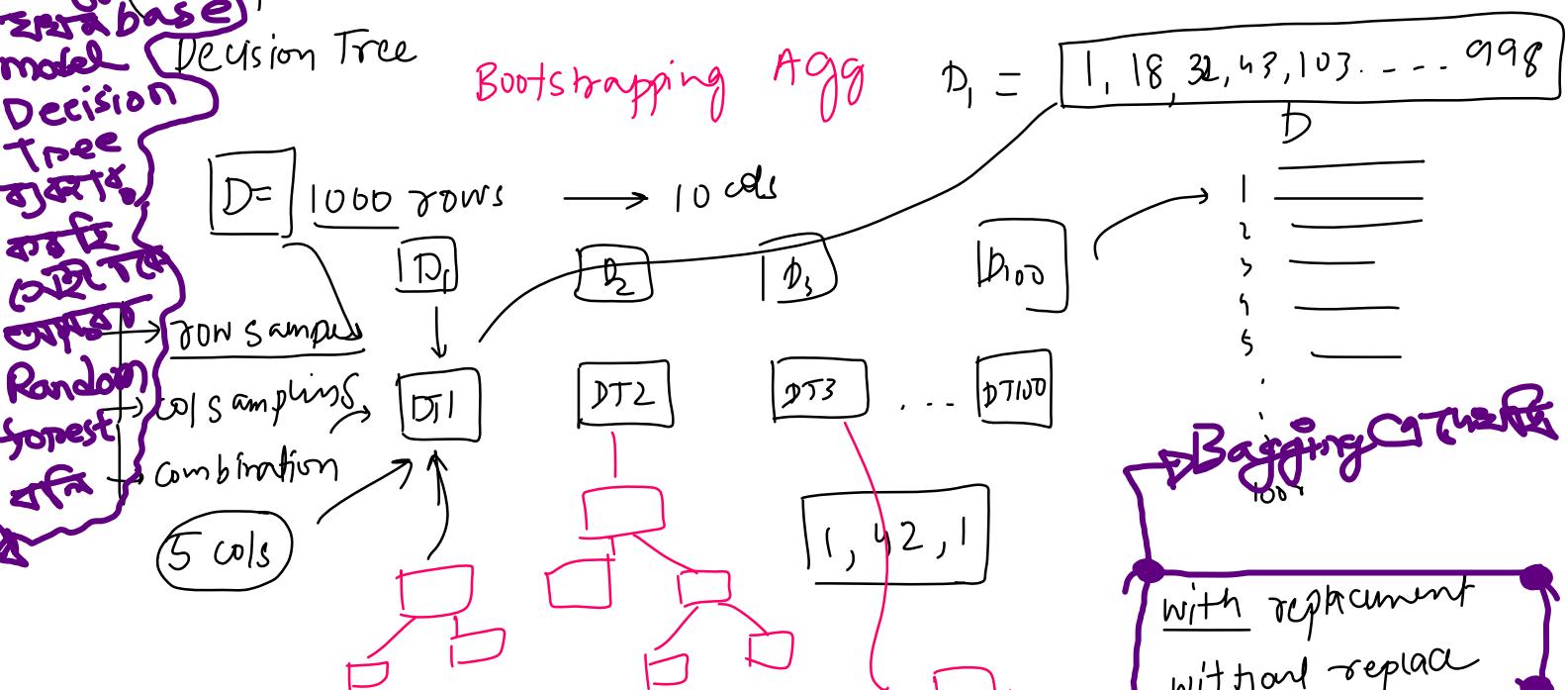


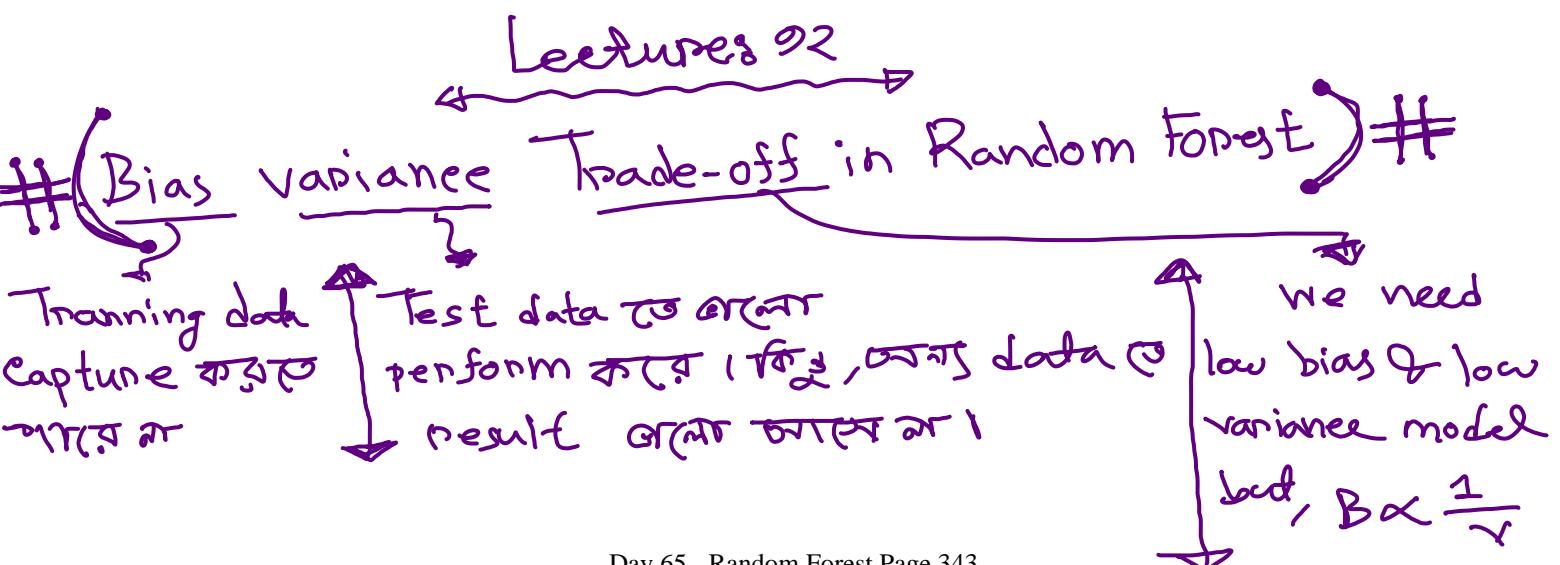
Trees, আমাকে Decision trees পাঠাই শব্দ।



বৈধ। কাব্য, Dataset পিছি, আকৃতি
 aggregation part এ, ট্রেইন ওস. গু দ্রোণি
 আমদে, দৈনিক রেজিস্ট্রেশন। Regression prob. থেকে mean
 calculate করছো।

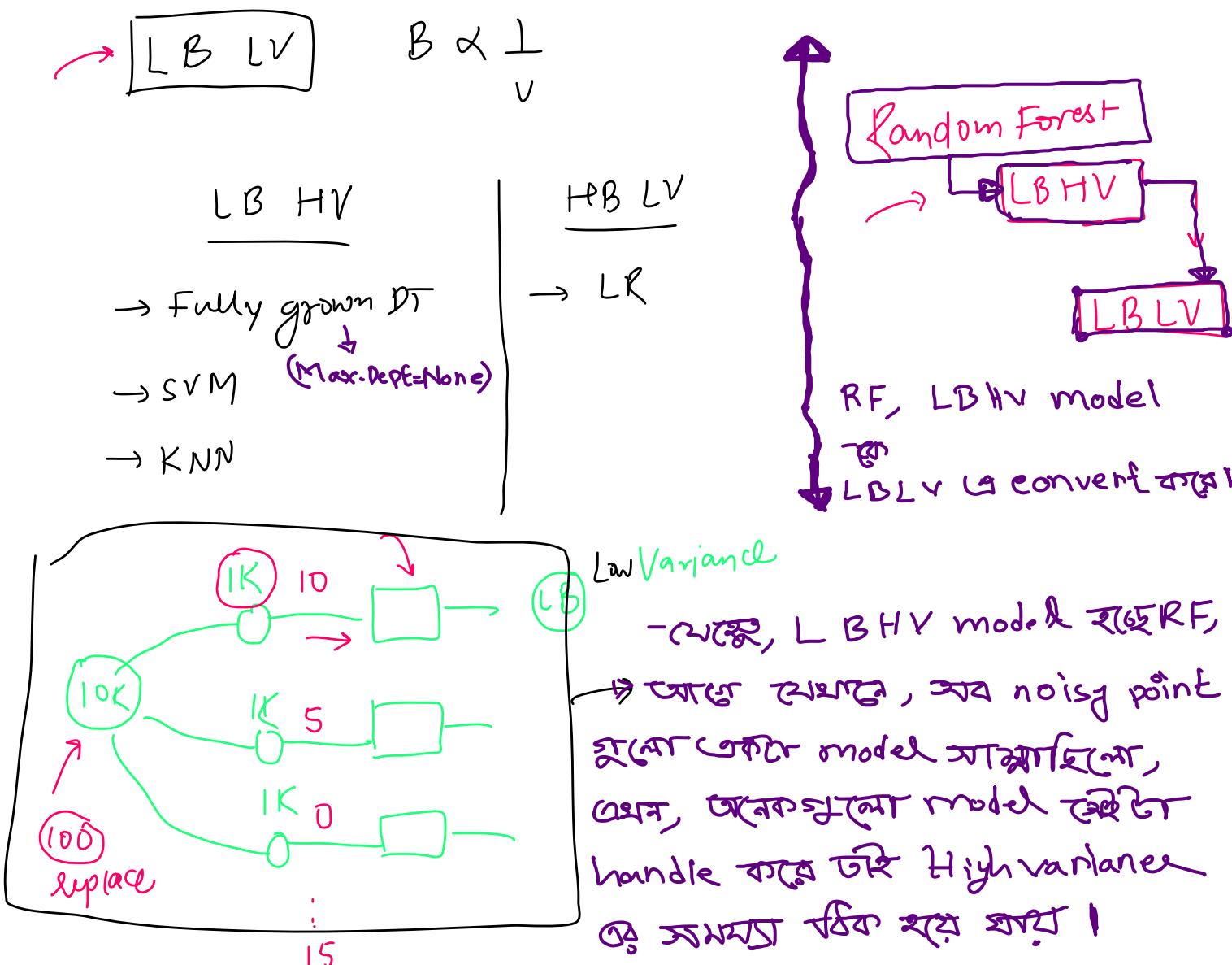
Demo

Monday, July 19, 2021 5:05 PM



Bias Variance and Random Forest

Tuesday, July 20, 2021 10:52 AM



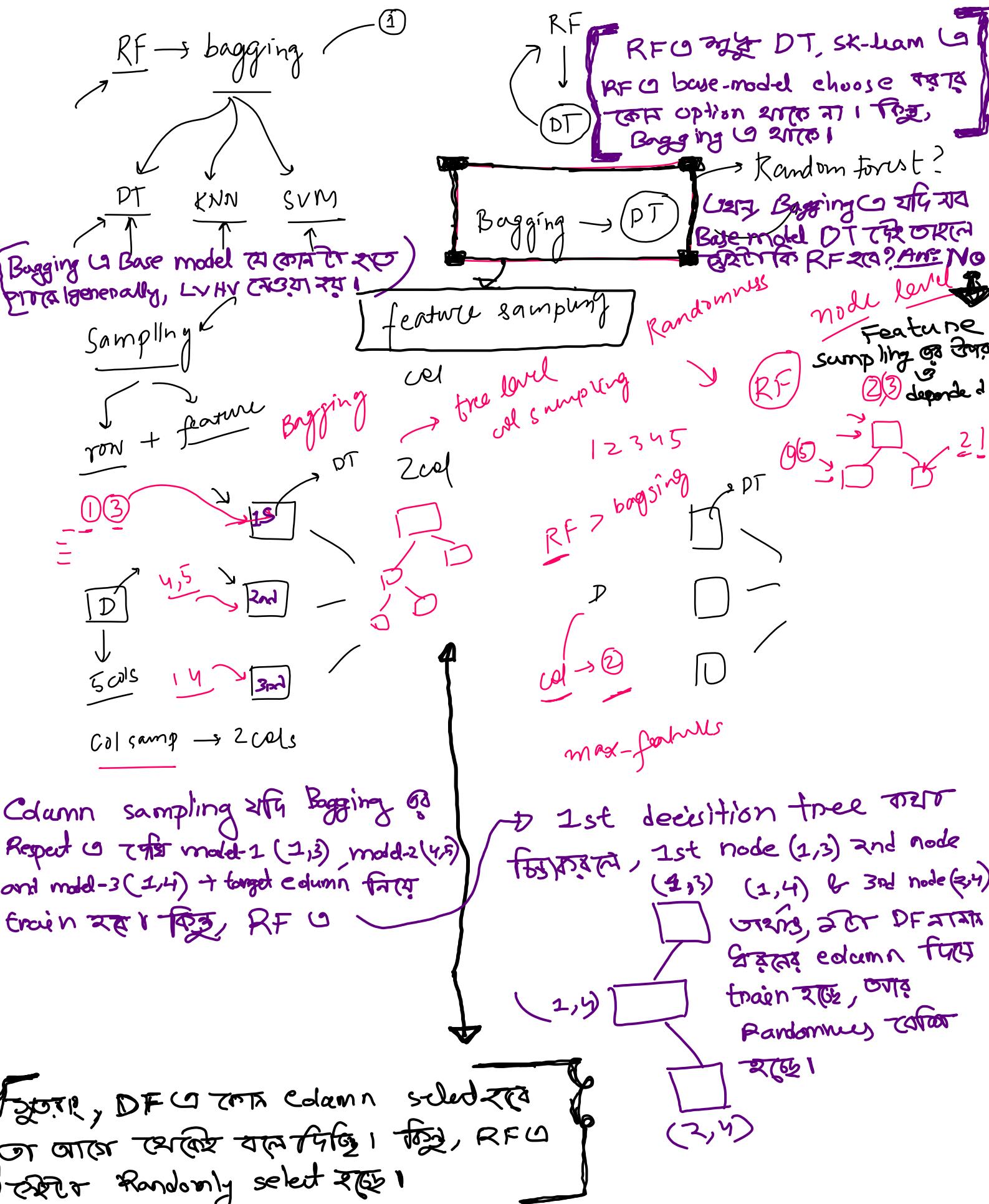
Interview Qs: Difference Between Bagging vs RF

Lecture: 93

* (Bagging vs Random Forest) *

Bagging Vs Random Forest

Monday, July 19, 2021 4:53 PM



Lecture # 94 (RF Parameters)

Hyperparameters

Monday, July 19, 2021 4:56 PM

Random Forest - Classifier

[Num-Estimators, Max-features, Bootstrap, Max-Samples] → Ques

Hyper-parameters से कौन से दिल्ले RF के tune करते चाहते होंगे।

Num-Estimators → कम करने DT बनाए।

Max-features → Input में total 5 column हैं, total 5 तो
व्यवहार करता हो चाहे असंभव।

Bootstrap → Randomly, Row or column select करते तकि किस?

Max-Sample → कम करने row select करते तकि ताकि असंभव।
→ (50 ~ 75)% data दिले जाने perform करें।

[उन्हें जितना important वाली हुई DT ले पाएंगे]

Random-Forest- Regressor → (ट्रेटमेंट में व्यवहृत)

Code Demo

Monday, July 19, 2021 5:05 PM

Lecture: 65

Hyper-parameters tuning in RF

with

GridSearchCV and
RandomSearchCV

[GridSearchCV] → data set ৰেখে কোন প্রক্ষেপ কৰি। যাইহ'ল,
time দেখিবিগুড়ো। ফল, Result, RandomSearchCV-তে অন্তর্ভুক্ত
আছে আবাস্থা possibility থাকু।

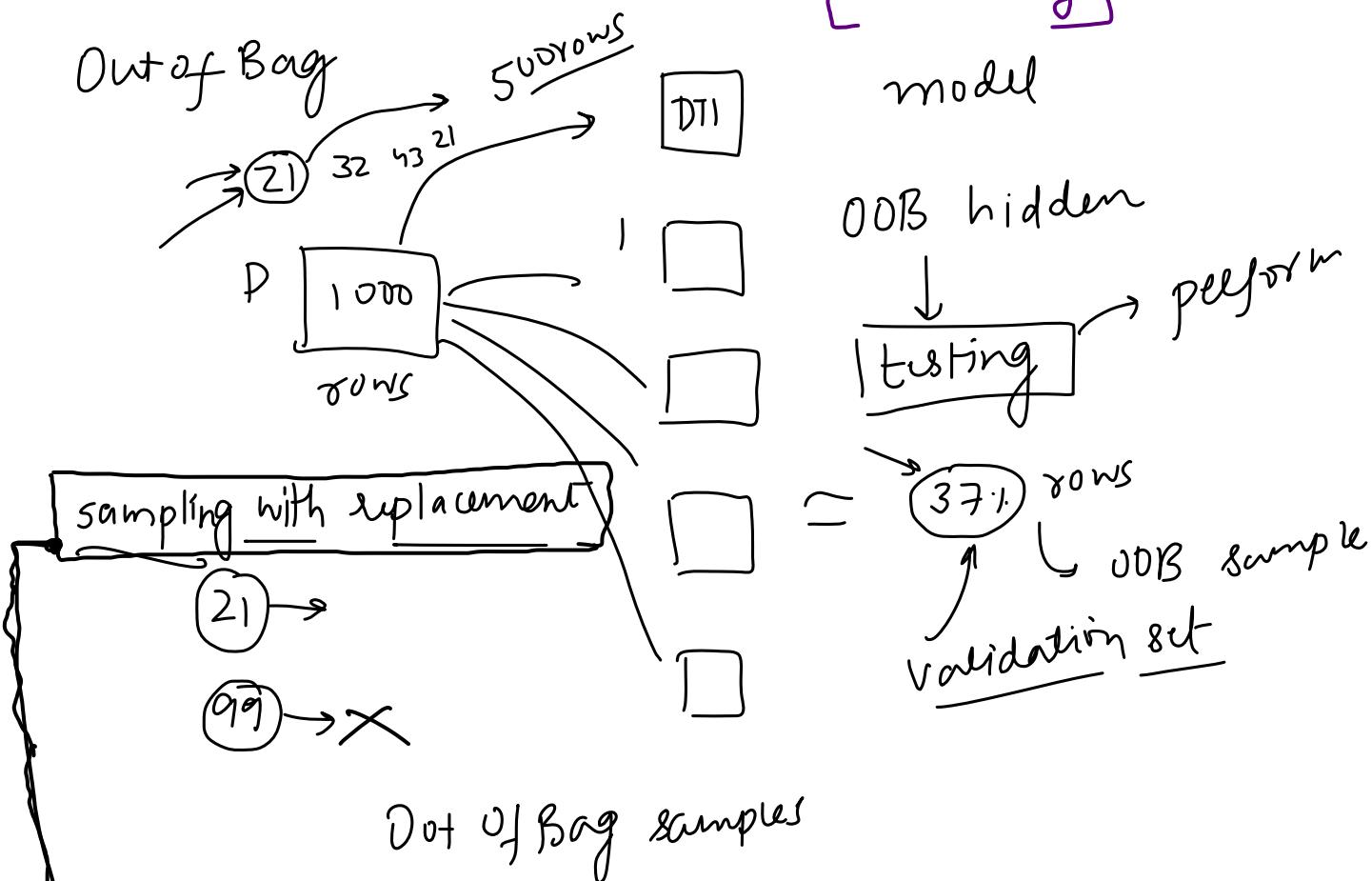
[RandomSearchCV] → data set যাচ্ছে কোন প্রক্ষেপ কৰি। যাইহ'ল,
Randomize, তাই result, GridSearchCV-তে কোন কোন প্রক্ষেপ
ৰেখায় আশঙ্কা থাকে।

Lecture 96 (OOB score/Evaluation)

OOB Evaluation

Monday, July 19, 2021 5:14 PM

[Out Of Bag]

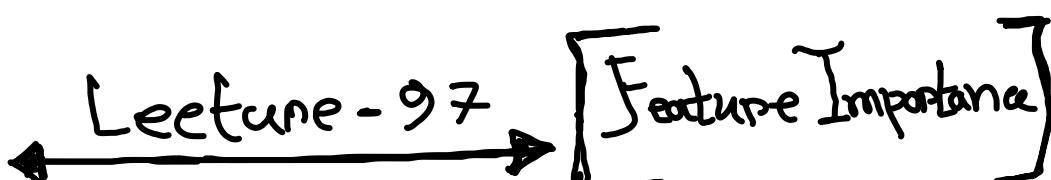


Sampling with replacement করলে, উপরের example

২১ অন্ত বাই select হয়েছে, ফিলি, ৭৭ অব্যাপ্ত select হয়নি, কোম্বা
base-estimator এবং এইধৃতিটি OOB sample কে OOB sample বলি।

কোরে OOB শুধুকে আশীর্বাদ collet করে testing করতে পারি।

Mathematically almost 37% OOB রয়ে থাকে।



Feature selection, এ সোন্মুহীন যোগাদৰ ব্যক্ত যদি ক্ষেত্ৰগুৰুত্বে feature
হৃচ্ছে এন্টেন্ট model converge কৰতে যেক time লাগবে। Feature
selection কোৱা ক্ষেত্ৰগুৰুত্বে important feature শুধু select কৰি।
অন্ত, Feature selection কোৱা ক্ষেত্ৰ technique আছে, যাই সৰু feature
importance কৰিব।

Feature Importance

Monday, July 19, 2021 4:56 PM

MNIST

label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel774	pixel775	pixel776	pixel777	pixel778	(784)
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	42000 images
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	



$28 \times 28 \rightarrow 784 \rightarrow$ Flatten apply 784×784 input

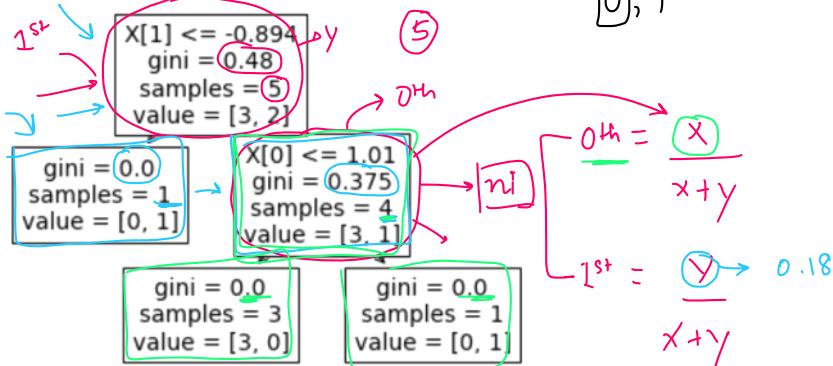
feature import
feature convert
pixel to column
Random forest
detected useful features
useful !

④ Random forest ने feature importance को इसके calculate करते जागे आगे याले DT
प्रक्रिया से विशेषज्ञता जारी रखे। \oplus 5

$N = 1^5$

$$\text{node importance } ni = \frac{N-t}{N} \left[\text{impurity} - \left(\frac{N-t-r}{N-t} \times \text{right-impurity} \right) - \left(\frac{N-t-l}{N-t} \times \text{left-impurity} \right) \right]$$

$$f_{ik} = \frac{\sum_{j \in \text{node split on feature } k} ni}{\sum_{j \in \text{all nodes}} ni}$$



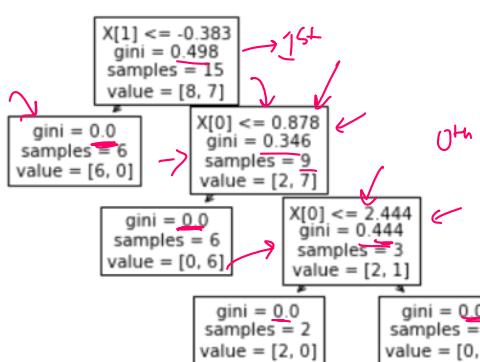
$$= \frac{5}{5} \left[0.48 - \frac{4}{5} \times 0.375 - \frac{1}{5} \times 0. \right] = Y \Rightarrow \left[0.48 - \frac{0.8 \times 0.375}{5} \right] \Rightarrow 0.48 - 0.30 = 0.18$$

$$\frac{4}{5} \left[0.375 \right] = X = 0.8 \times 0.375 = 0.30 = X$$

$$0^{th} = \frac{0.3}{0.3 + 0.18} = 0.625$$

(feature importance of 0th column) $D_{0th} = \frac{0.3}{0.3 + 0.18} = 0.625$ ↗ ni of 0th column

" " " 1st column) $I_{1st} = \frac{0.18}{0.3 + 0.18} = 0.375$ ↗ ni of 1st column



0th Node

$$\frac{15}{15} \left[0.49 - \frac{9}{15} \times 0.34 \right] = 0.290$$

$$\begin{aligned} & \frac{9}{15} \left[0.34 - \frac{3}{15} \times 0.44 \right] = 0.118 \\ & \frac{3}{15} \left[0.44 \right] = 0.088 \end{aligned}$$

$$f_i[0] = \frac{0.11 + 0.08}{0.29 + 0.11 + 0.08} = 0.48 \\ = 0.39$$

$$f_i[1] = \frac{0.29}{0.48} = 0.60$$

Feature importance in Random Forest ↗

Random Forest ইতে, একাধিক DT তেই একই column টো
feature importance কোধিক যাবু কালাকালে সংযুক্ত শৰু
Average করে আসবো value → f(i) ।

- High Cardinality feature হিবেন্ট, অম্বুন্ডা, decision tree ই random forest ও চেয়ে
`sklearn.inspection.permutation_importance()` মাত্র
এটা fn আছে, এইটোই ব্যবহার করতে পার(গো)

Extra Trees

Monday, July 19, 2021 4:57 PM