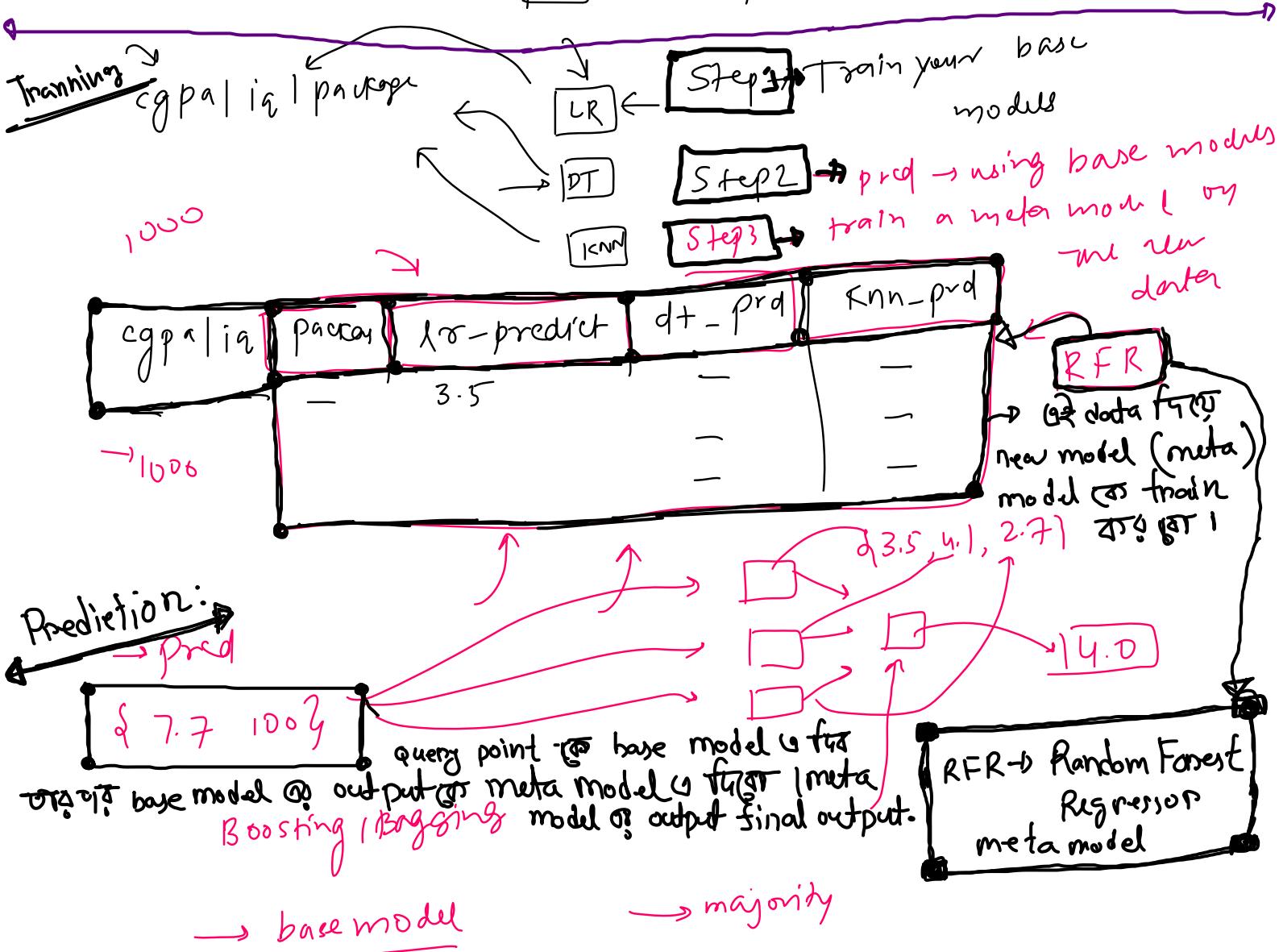
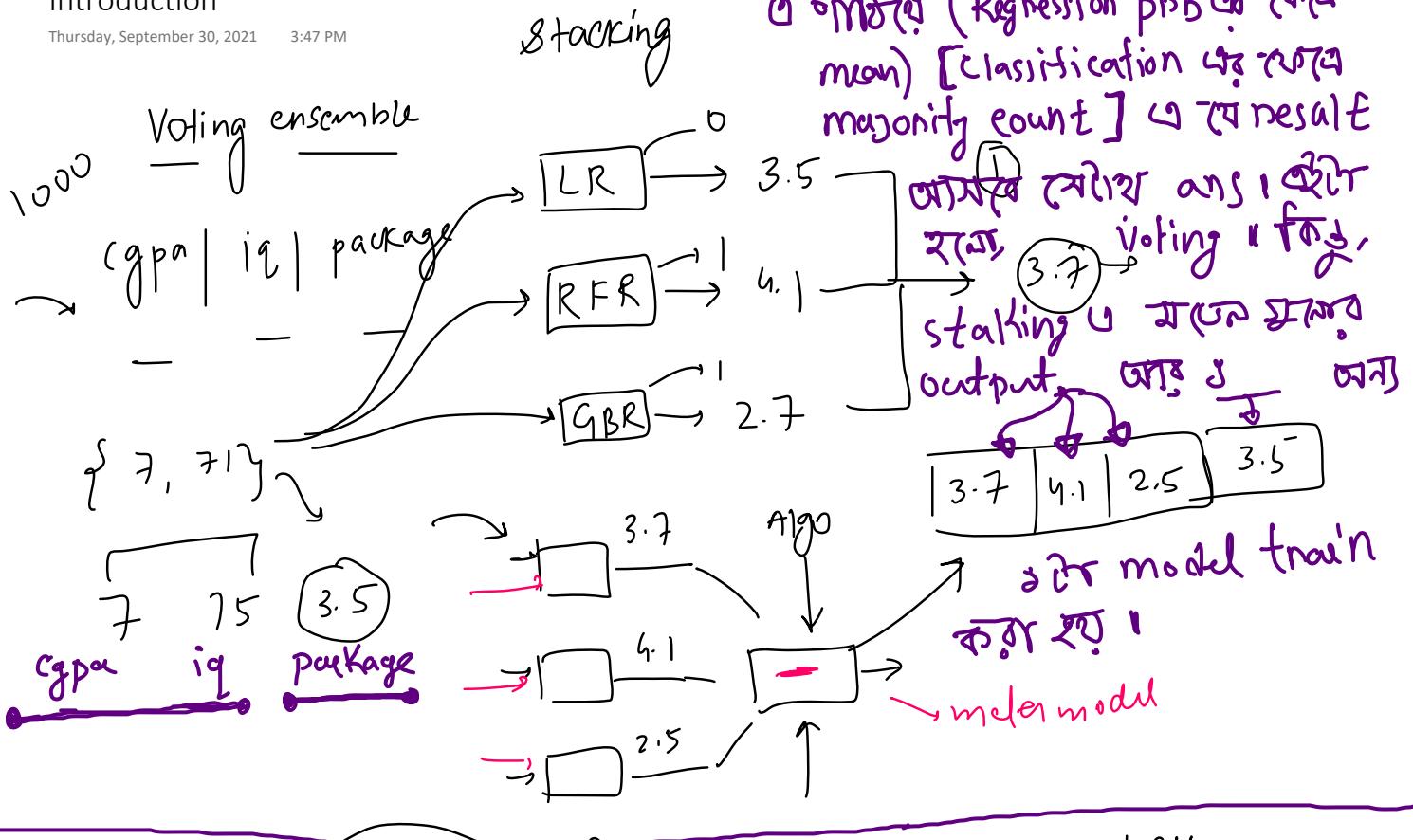


Lecture 109 (Staking)

Introduction

Thursday, September 30, 2021 3:47 PM



→ training

ପାଇଁ
Bagging, Boosting

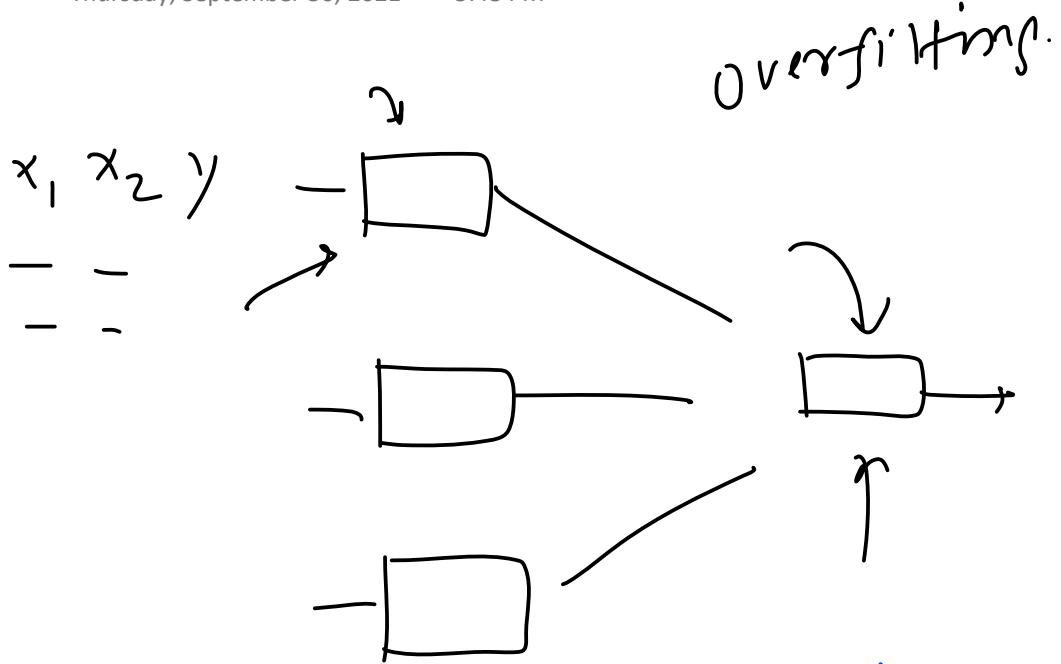
i) Bagging & Boosting ଏ Base model କେମ୍ବାକୁ ।
ଫୁଲ୍, stacking ଅତି ବୋଟିଙ୍ ଏ base model ଆବଶ୍ୟକ ।

ii) → Bagging ଏ majority count ଅତି ମେନ୍ଦିରି ।
→ Boosting ଏ weighted sum ହେବୁ କାହିଁ ।
ଫୁଲ୍, stacking ଏ କ୍ଷେତ୍ର ମୋଡ଼୍ଳୁ ଫ୍ରେଣ୍ କରି ଏକ ଫିଲ୍ଡିଙ୍ ପ୍ରେଡିକ୍ଷନ୍
କରୁଛି ।

Problem with stacking

Problem

Thursday, September 30, 2021 3:48 PM



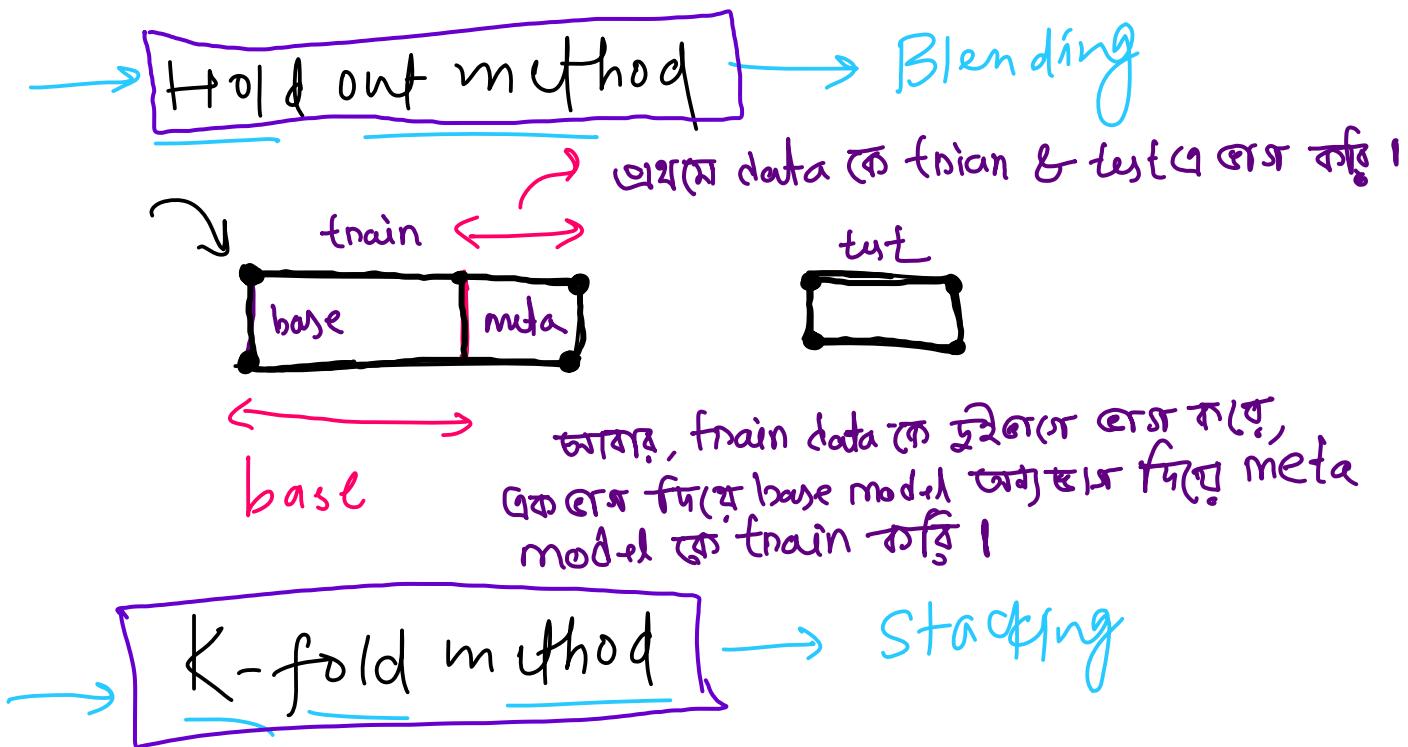
जानें कि same data का ट्रॉइंग training करते, predict करते ही पत्ते जानकारी में overfit करते हैं। Base mode या LBHV रूप से overfitting की possibility जानकारी बहुत अचूक होती है।

Solution of Overfitting

- i Hold out method or bending.
- ii K-fold method or stacking.

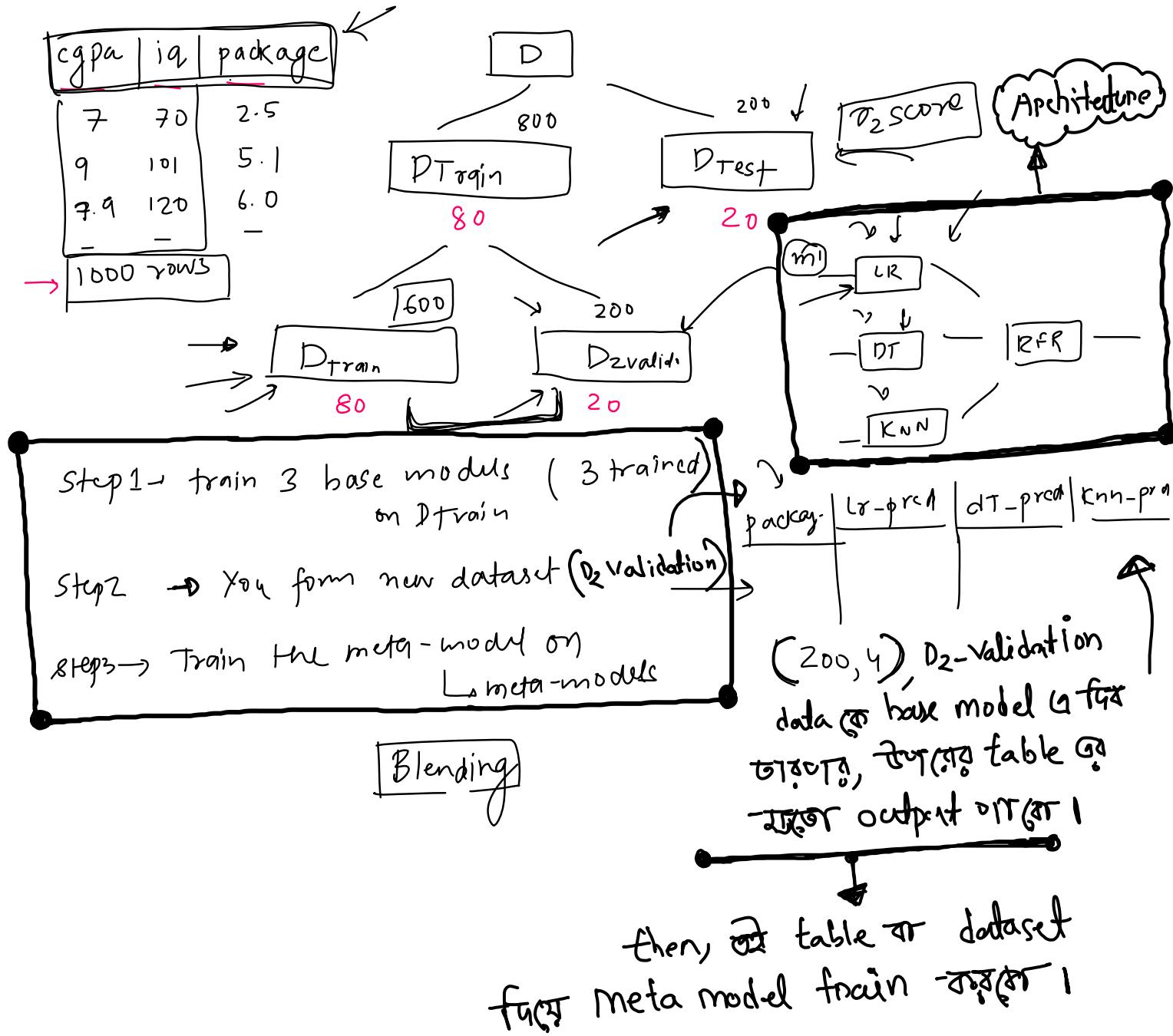
Solutions

Thursday, September 30, 2021 3:48 PM



Hold Out Approach - Blending

Thursday, September 30, 2021 3:48 PM



- ④ D_{test} data বাছতে, মেটা মডেল দিয়ে আবশ্য টেস্টিং কোর্ট
- কোর্ট পাইগুলি । ④

K Fold Approach - Stacking

Thursday, September 30, 2021 3:49 PM

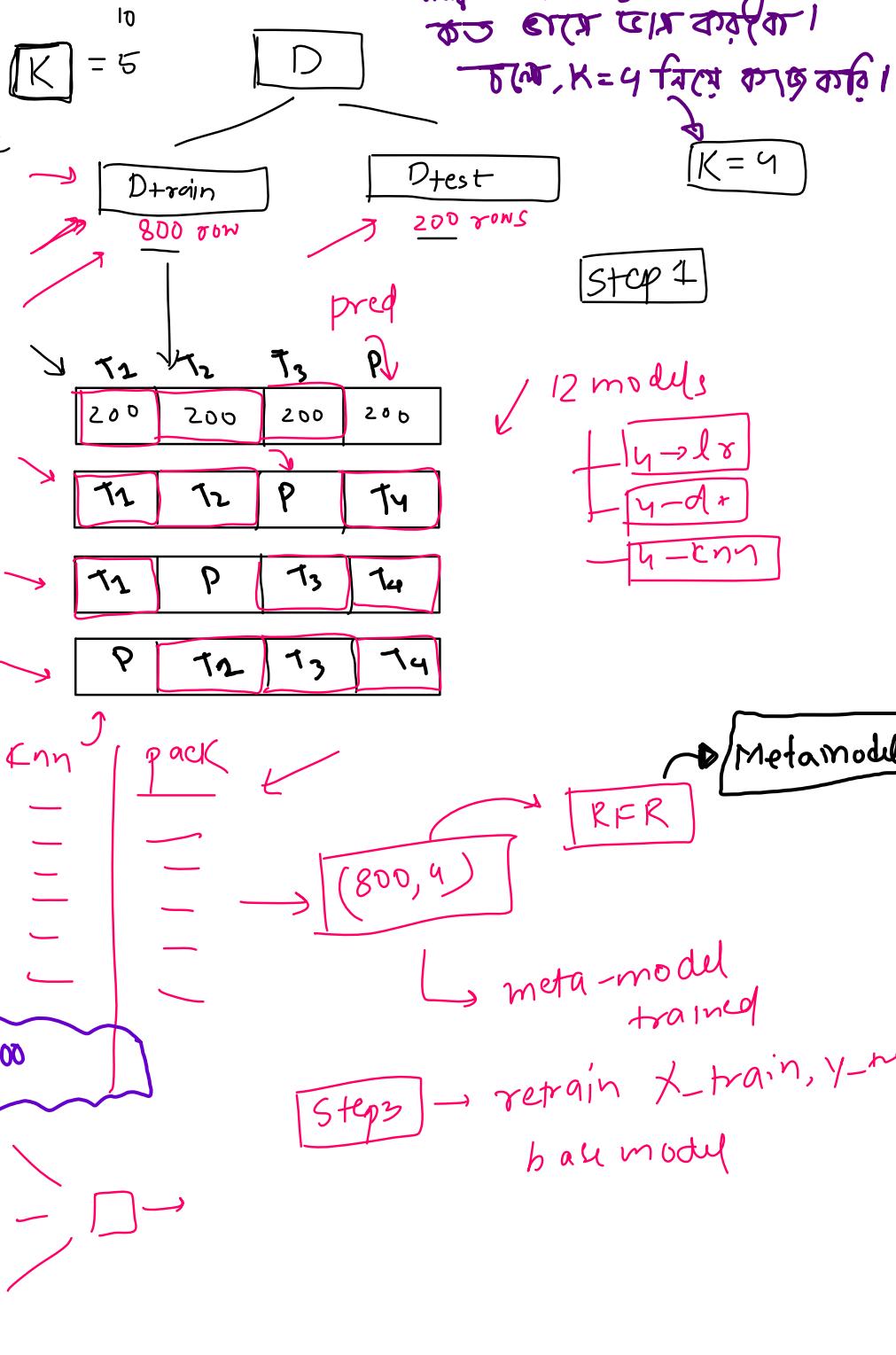
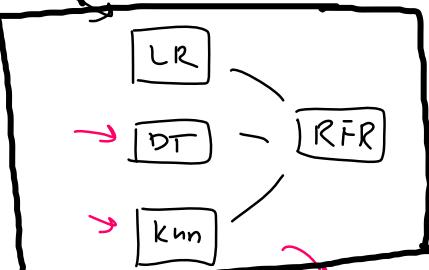
Stacking

(K-fold), Generally, K=5, K মাত্র
আসুন আমরেরই training data কে
কো টেস্ট আবশ্যিক করা।
চলো, K=4 বিষ্ণু কীভাবে।

	cgpa	iq	package
7	70	2.5	
9	101	5.1	
7.9	120	6.0	

Architecture

1000 rows

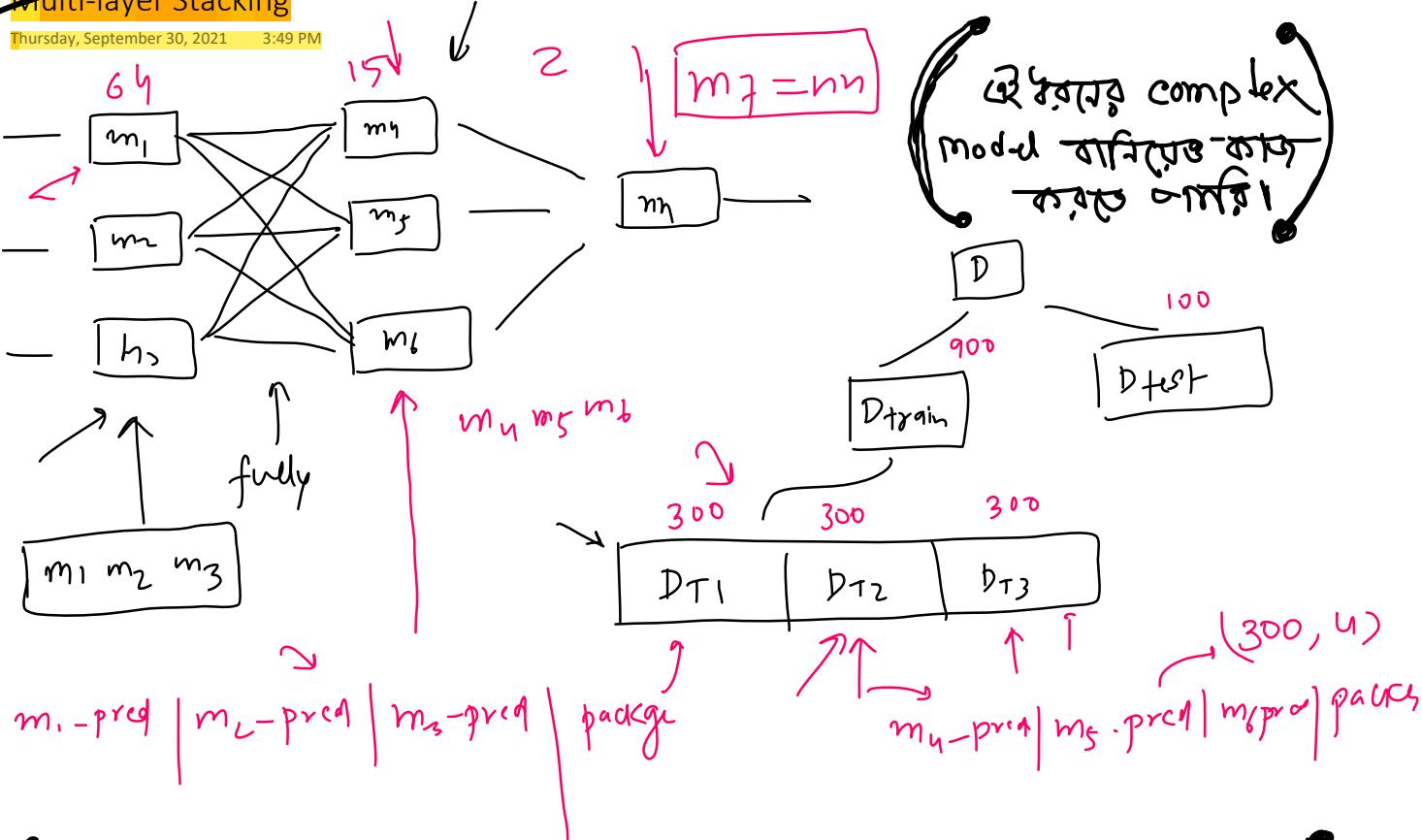


আসুন D_{train} data কে 4 অংশে ভাগ করি, LR কে 1st রাখো $T_1 T_2 T_3$ dataset
ফিল্ডে train করো পৰিষে PRediction কো হ'ল In-Predit (200×1) value কো হ'ল ই।
যাবলি 3 অংশ data কো হ'ল কুলে total. LR-Predit মানে (800×1) 1 simillarly,
DT, KNN অৱৰ, (4×3) = 12 বিষ্ণু model train কো হ'ল। এখনোপৰু লু-pred, DT-Pred,
KNN-Pred দ্বি মাথে actual y মান কো হ'ল (800×4) meta model কো train
কুলো। D-test (200×1) এ dataset ফিল্ডে model কো test কুলো।

এখন, Prediction দ্বি মাথে total 12 বিষ্ণু base model থাকবে কিন্তু, আমোগ এদের পৰি কো
PRediction ও পৰিষের কুলো। Base model মানোঁ আজু কুলো কুলো D-train &

Multi-layer Stacking

Thursday, September 30, 2021 3:49 PM



এই স্টেকিং complex model ব্যবহার করে নিয়ে রাতে কাজ করতে পারি। SK-learn-এ implemented কোন API আছে না।

sklearn.ensemble.StackingClassifier()

Parameters

estimators → (base model)

final_estimator → (meta model)

cv → Cross Validation

stack-method → (output of classification in probability or 1,0 etc.)

Pass-through → meta model train করাটি যদিয়ে আমরা চাই
নেটুল এনেডেডেড (base model) থেকে ব্যবহার করেছি। Pass-through
= true করলে, main (রেখা), column (e.g., iq) স্টেপগুলোও input
দিয়ে রেখামডেল এ এসে।

Demo

Thursday, September 30, 2021 3:49 PM

Lecture: 110
Agglomerative Hierarchical Clustering

(ଆମେ ଆମ୍ବିକାନ୍ତର୍ଗତିରେ | KNN ଏବଂ fpg୍ରୁ | imputation କରିବାକାହିଁ)

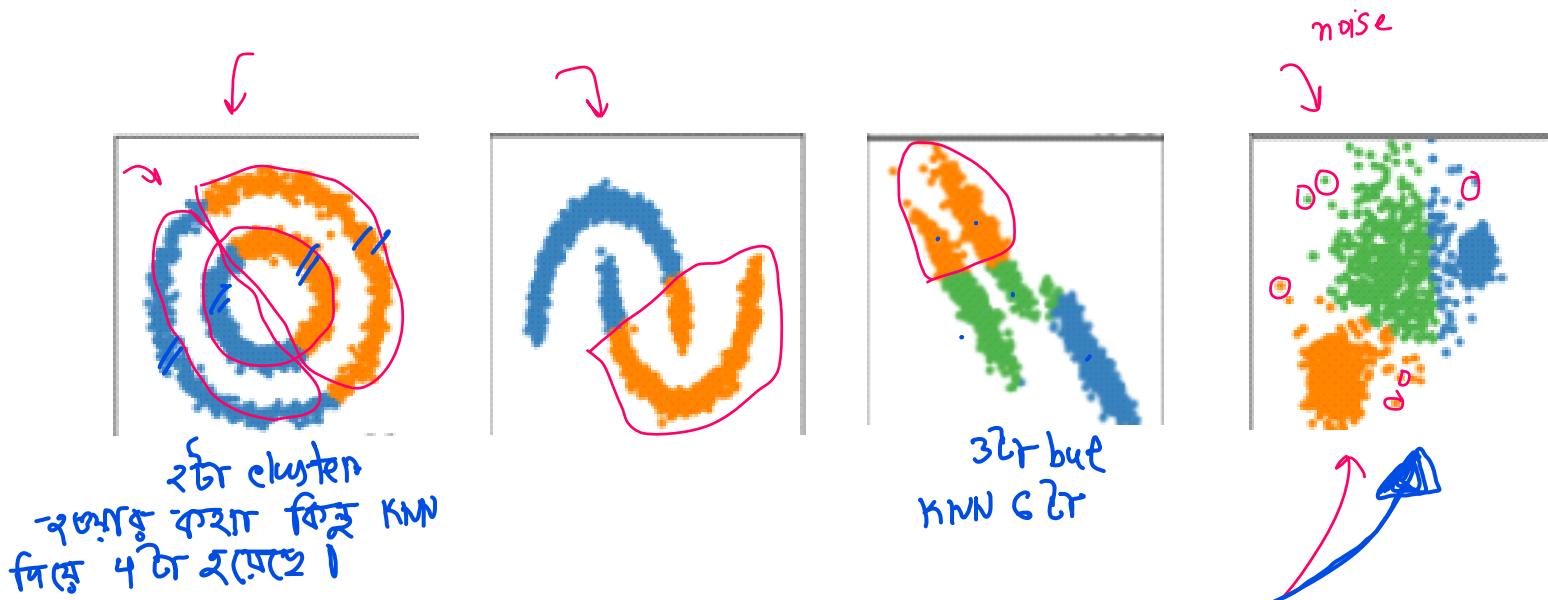
Need of Other Clustering Methods

Saturday, November 6, 2021 7:38 AM

Three colored shapes are arranged vertically. The bottom shape is orange, the middle shape is blue, and the top shape is green. Each shape is enclosed in a red oval.

KNN- centroid দুরত্বে distance calculate করে।

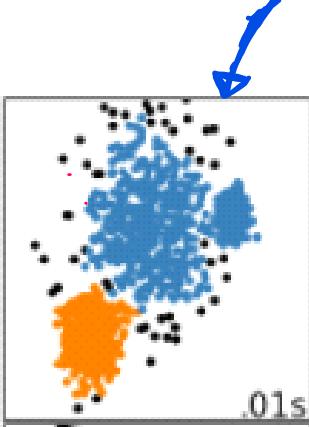
cluster টির কেন্দ্র। তাই, একটি
বর সীপেল type এর clusters এর
অস্য KNN গুরুত্ব কাছে রয়ে। অন্যদিক
মিঠু দুরত্বে complex clusters ঘো অস্য KNN গুরুত্ব
-কাছে রয়ে গা।



2 More clustering methods:

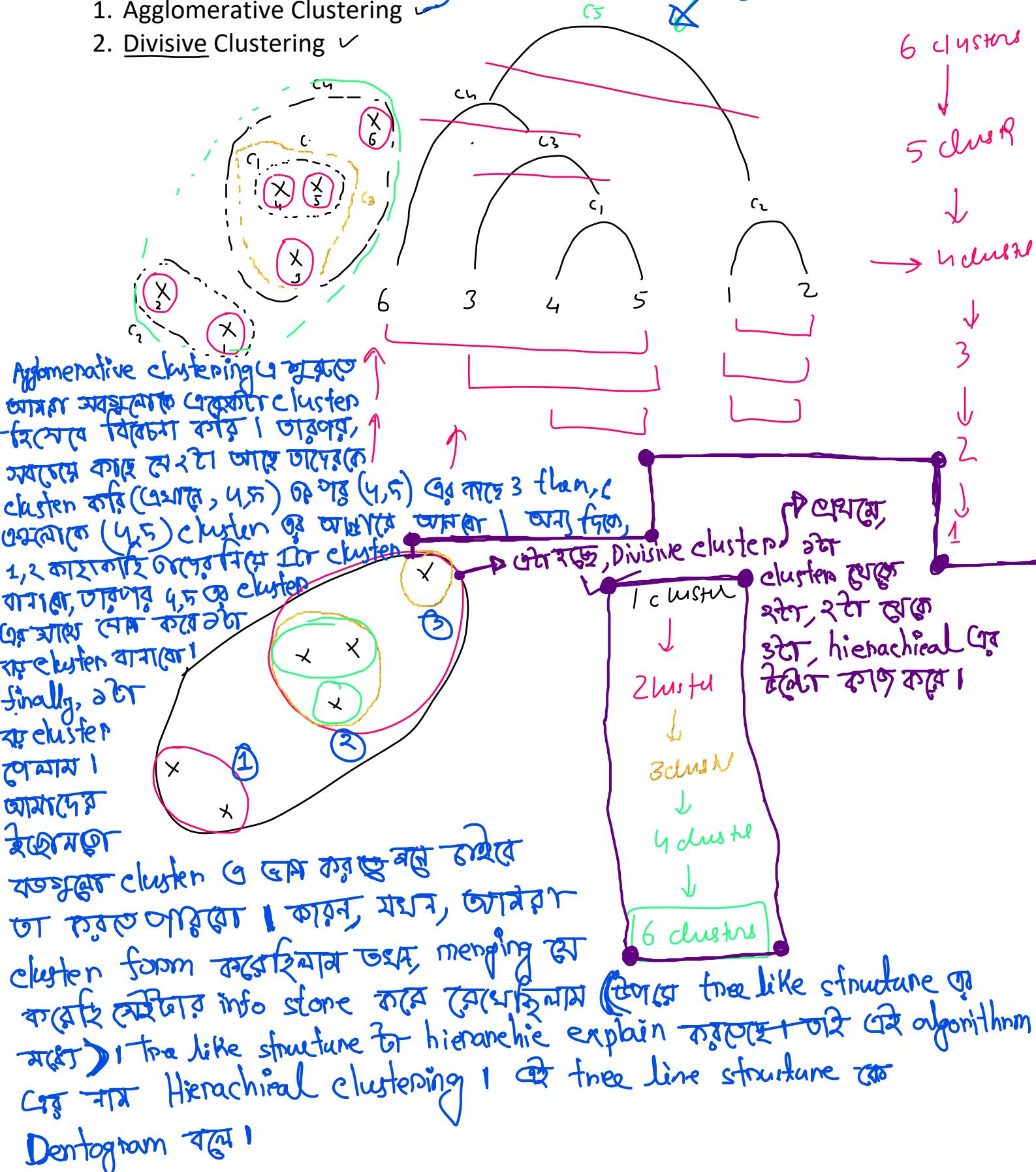
1. Hierarchical Clustering
 2. Density Based Clustering

SCANN সেপ্টেম্বরে cluster করা হয়। noise আছে। যেখানে KNN-কার্যকৃত করুণ
পাইবে না। DBSCAN, Density Based
clustering Algorithm like DBSCAN
noise-এর outliers-কার্যকৃত করুণ পাইবে।
যাতে ফালি সেপ্টেম্বরে ইবিজি DBSCAN দিয়ে
বিচুরি মতে দেখাবে।



Types of Hierarchical Clustering:

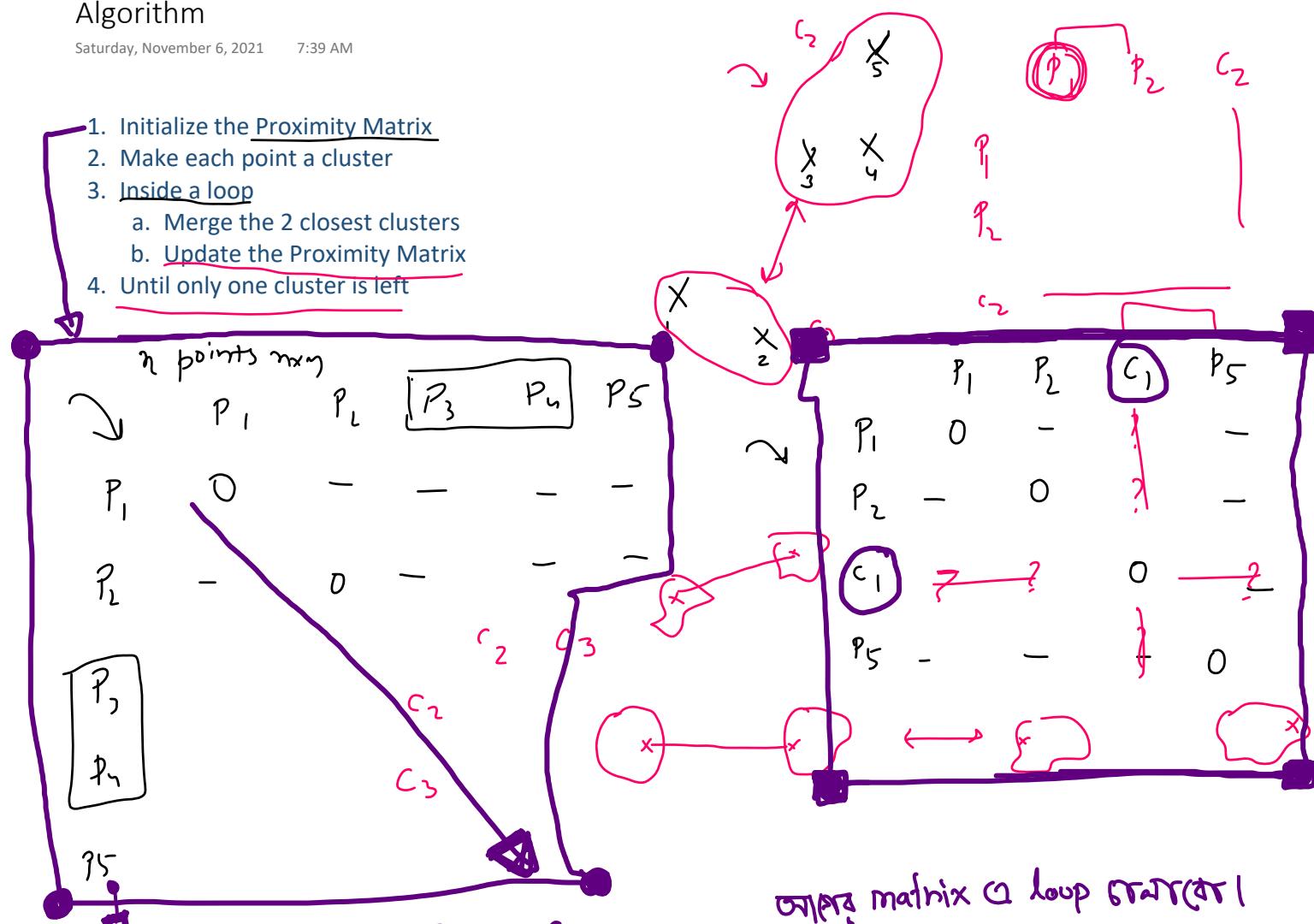
1. Agglomerative Cluster
 2. Divisive Clustering ✓



Algorithm

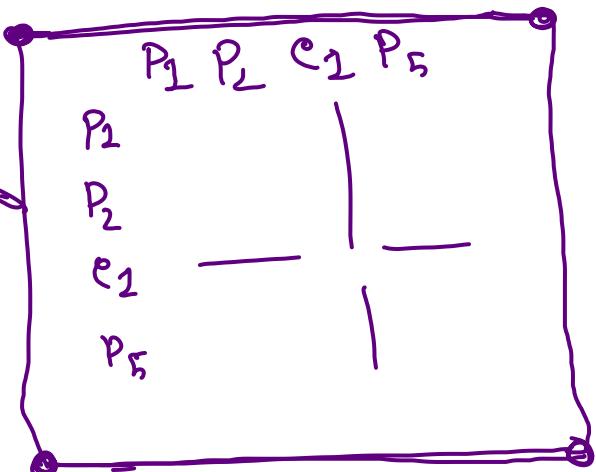
Saturday, November 6, 2021 7:39 AM

1. Initialize the Proximity Matrix
2. Make each point a cluster
3. Inside a loop
 - a. Merge the 2 closest clusters
 - b. Update the Proximity Matrix
4. Until only one cluster is left

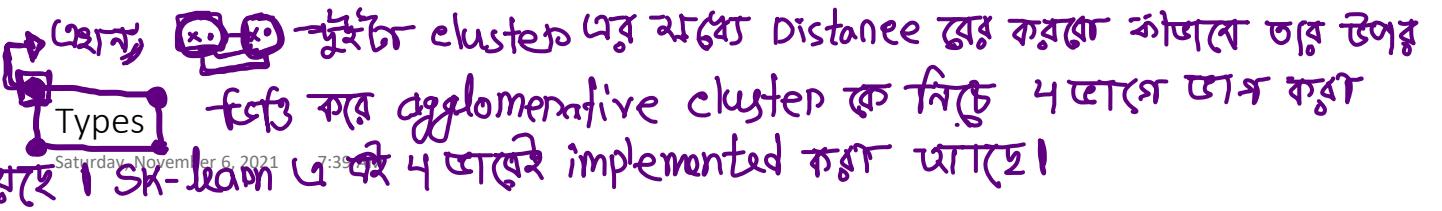


যামাদুর কালু এটা শব্দের আছে, তার অঙ্গে এই matrix রয়ে। যদি n থাকে তার তালু n×n এই matrix রয়ে। সম্পাদন, P₁ P₁ এর মধ্যকার দূরত্ব, এবং P₁ এর সাথে P₂ টির value শুধু matrix এর মধ্যে insert করতে হবে। P₁ P₁, P₂ P₂ গিয়ে সাথে নিজের distance 0। তবে, matrix এর diagonal elements কুন্ত হবে।

আপনি matrix ও loop চেয়ে।
দেখো কেন point কে কৈ
সবচেয়ে কাছে। উদাহরণ example
এ P₃ P₅, তবে কেনেকে নিয়ে
এ নাকে নতুন cluster বানিয়ে
পুরুষ proximity matrix
বানাবে। তাহলে, P₃ P₄ কে
C₁ দিয়ে identify করুন,

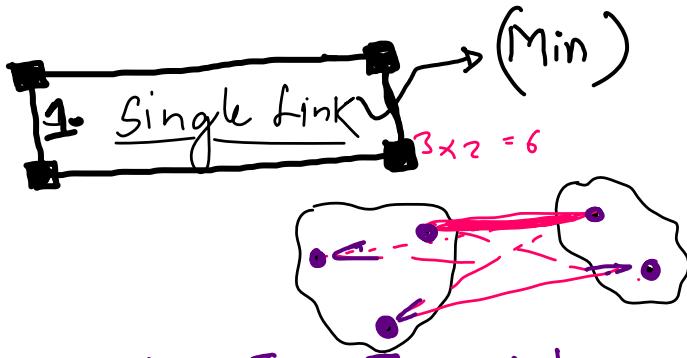


কেবল করতে পাঠে যতক্ষণ
না পর্যন্ত matrix তে size (1×1)
এটা cluster এর একেবারে না
হাতু।

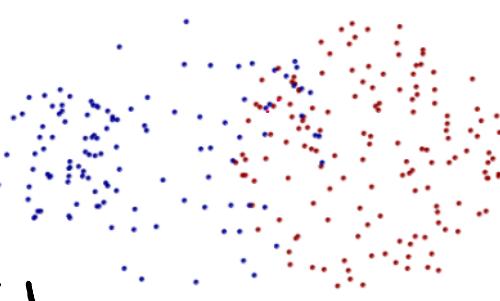
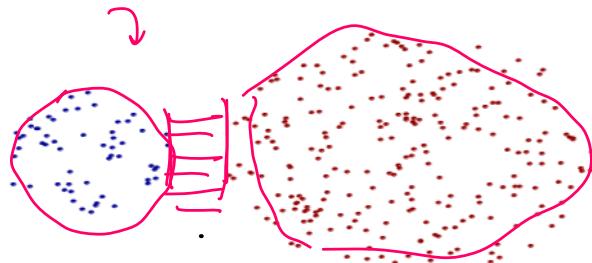
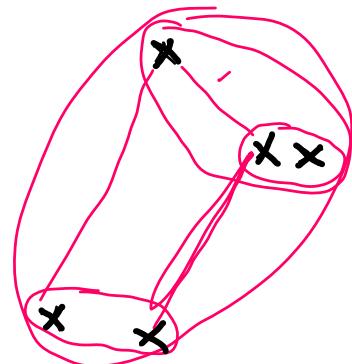


Types of Agglomerative Clustering

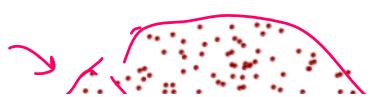
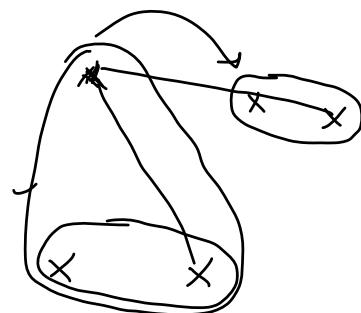
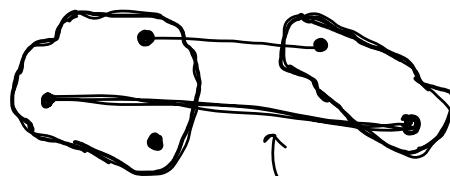
1. Min (Single-link)
 2. Max (Complete Link)
 3. Average
 4. Ward



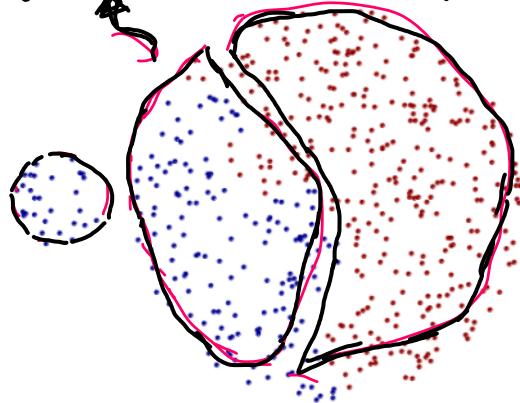
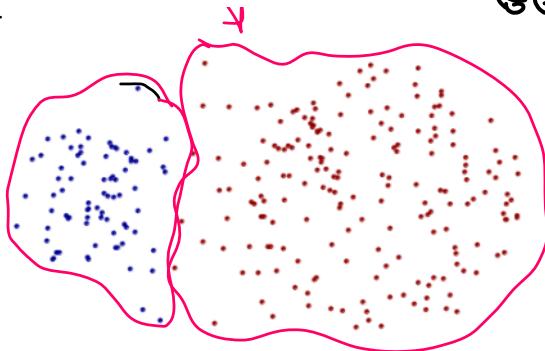
2 टो cluster पर $[3 \times 2 = 6]$ घेरे distance येणे काळा पाहा ।
एधन, single link पर 6 टो distance तरी मर्टीज गवऱ्याशे अरीता करा विचार करा ।



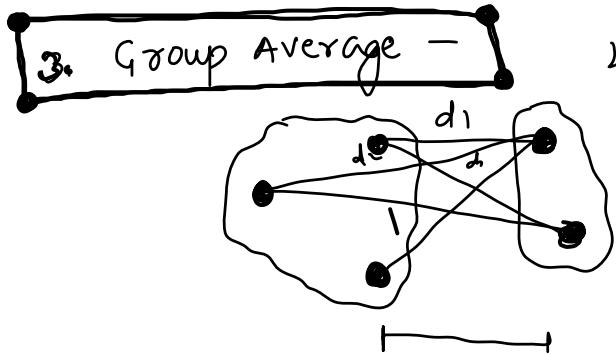
Outliers ଯାକଣ୍ଟି single Link କାହିଁ ରହୁଥିବା ।



एटा cluster द्वाटे अन्य एटा cluster वडा । वडीले
जुऱ्हे मध्ये रद्द टुकडी शहायेहो । Complete link
41

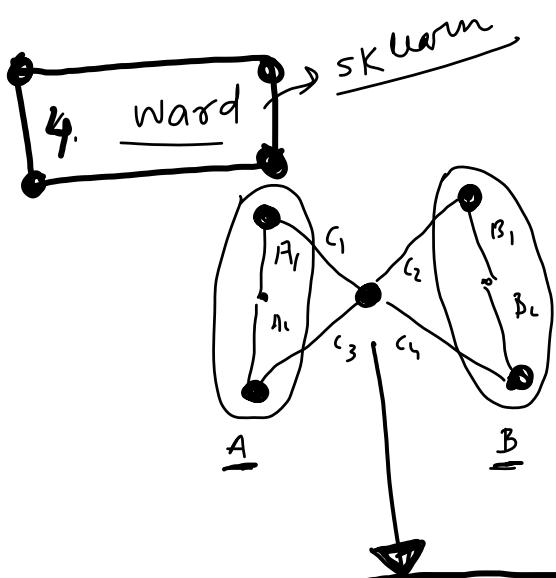


outliers थांगाने handle रस्ता automatically ।



$$\frac{d_1 + d_2 + d_3 + \dots + d_4}{2 \times 3}$$

$3 \times 2 = 6$ दूरी distance
वर्ष काळा average
distance वर्ष काळा ।



$$dist = \frac{c_1^2 + c_2^2 + c_3^2 + c_4^2 - A_1^2 - A_2^2 - B_1^2 - B_2^2}{4 \times 3}$$

Variance minimize

sk-learn (default रस्ता),
ward

Centroid रेष्ट काढी । दुसरी cluster
ऐसु एकल data point घेके ठारूणार,
centroid घेके दो एकल cluster घेण
point समान दूरी, e_1, e_2, e_3, e_4 ,
बर्ग करून योग काढी । ठारूणार एक
एकल cluster घेण्याची centroid रेष्ट
करून अतिरिक्त point घेकू दूरी
रेष्ट करून subtract काढी ।

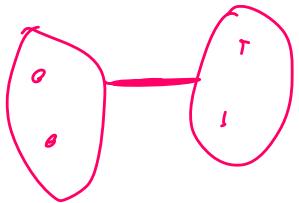
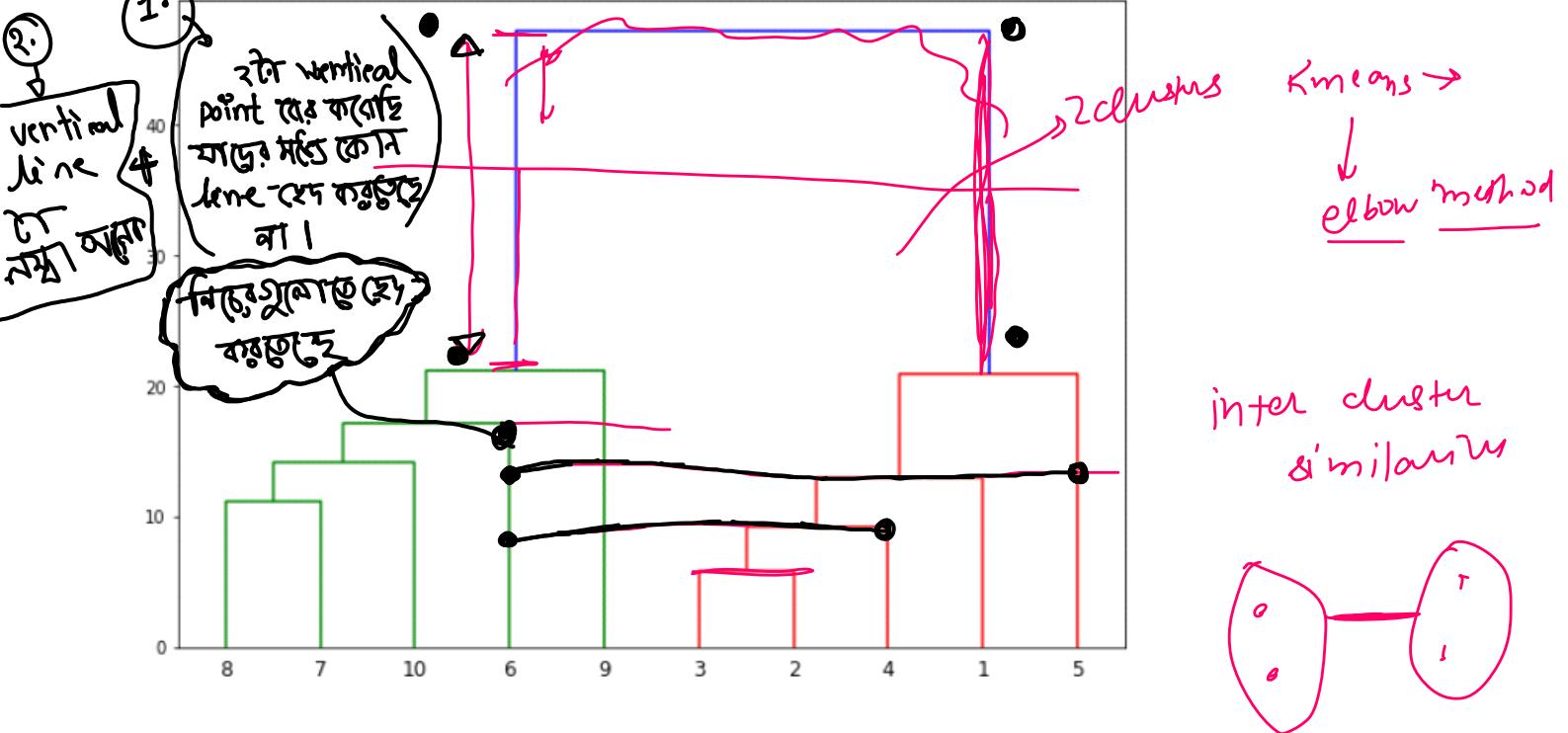
কতজনো cluster বানাতে হবে তাইবা সুবাদে কাণ্ডে ? ?

How to find the ideal number of clusters

K-mean Elbow method - এব্যাবহৃত করেছি

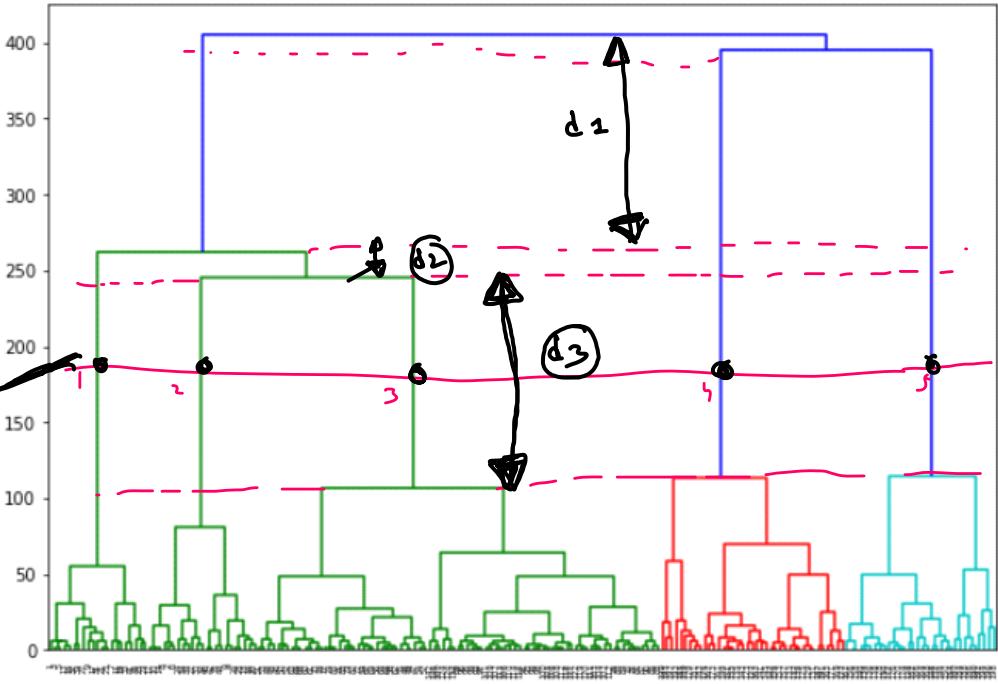
এখন আমরা Dendrogram ব্যবহার করবো ।

2টা সাধেই খেয়াল কুণ্ডলো



Customer Dendograms

Cutting point



d_3 টুর distance
সংজোড়ে ভেঙ্গি

সাধারণত ফ্রেন
সহজেই ভেঙ্গি

সহজেই ভেঙ্গি

- কেবি ।

আবু লাইফি ৫ জাপানীয় cut

করে, তাই cluster number হবে ৫ ।

ওম্বৰ, vertical নাইন
intra cluster similarity
mean করে ।

Hyperparameter

Saturday, November 6, 2021 7:40 AM

In sk-learn,
sklearn.cluster.AgglomerativeClustering()

For hyperparameters see the sk-learn
Documentation.

Code Example

Saturday, November 6, 2021 7:40 AM

Benefits/Limitations

Saturday, November 6, 2021

7:39 AM

Benefits

1) Widely applicable

2) Dendogram

