

depth = 2
depth - large
- overfit

→ 8 → 32
→ 64, 128
✓

64 conditions

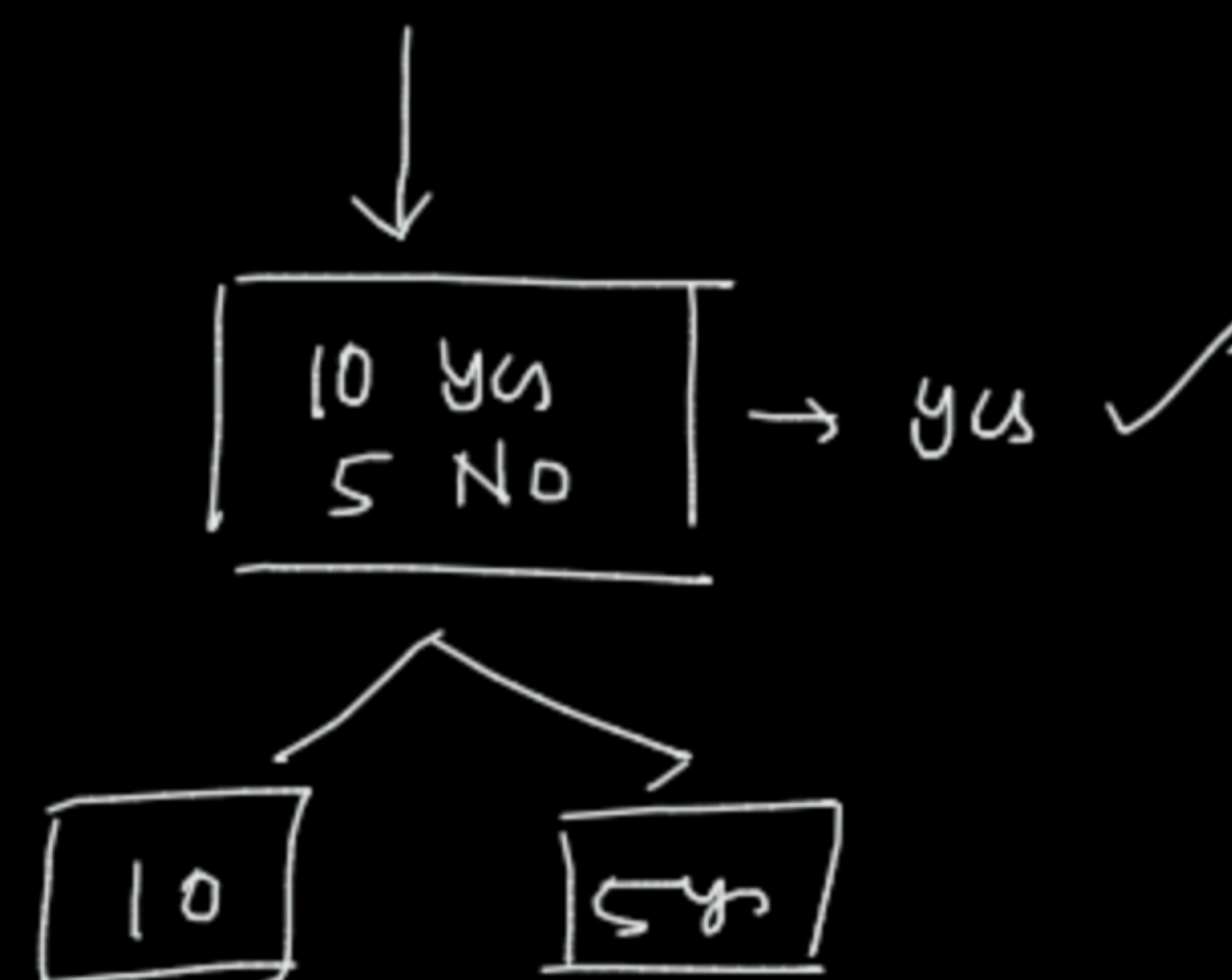
Training Data → Extremely well

Testing Data → Badly

→ 5 ppl
3 Yes
2 No → Yes-✓

Stopping Criteria

- 1 max-depth = 4 ✓
- 2 max-leaves = 10 ←
- 3 min-samples = 10



1 Stopping Criteria

- 2 Entropy or Gini Impurity
- ↓
 C5.0
- ↳ CART

Classification

Entropy & IG $\rightarrow \checkmark$ \leftarrow Splitting done
Gini Impurity $\rightarrow \checkmark$

Regression

MSE $\rightarrow \leftarrow$ Splitting done



Opera

File

Edit

View

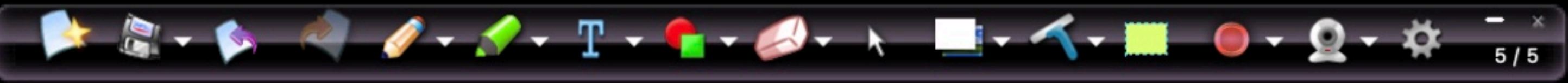
History

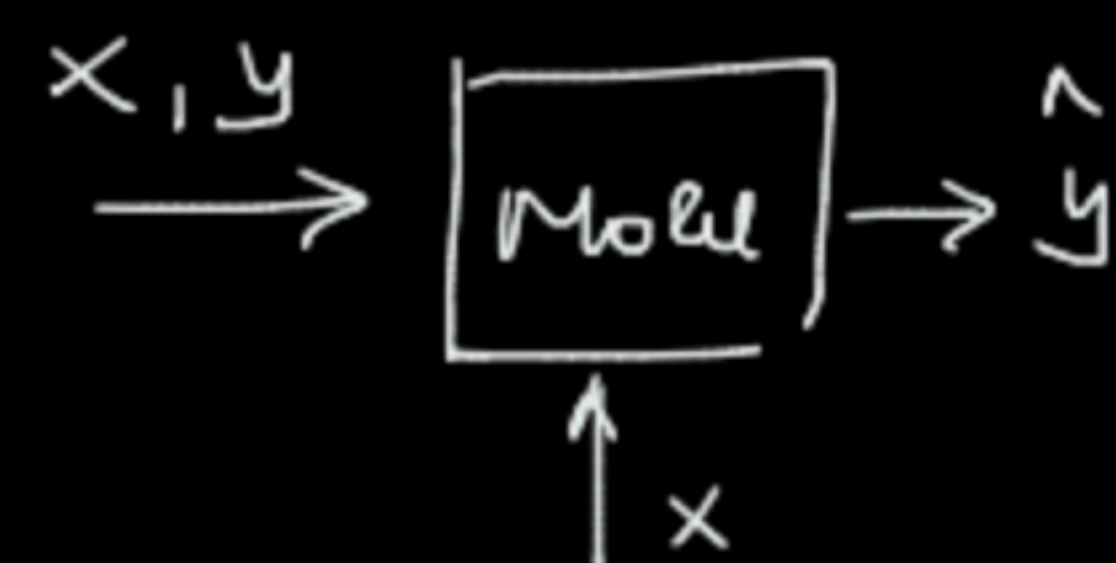
Bookmarks

Developer

Window

Help



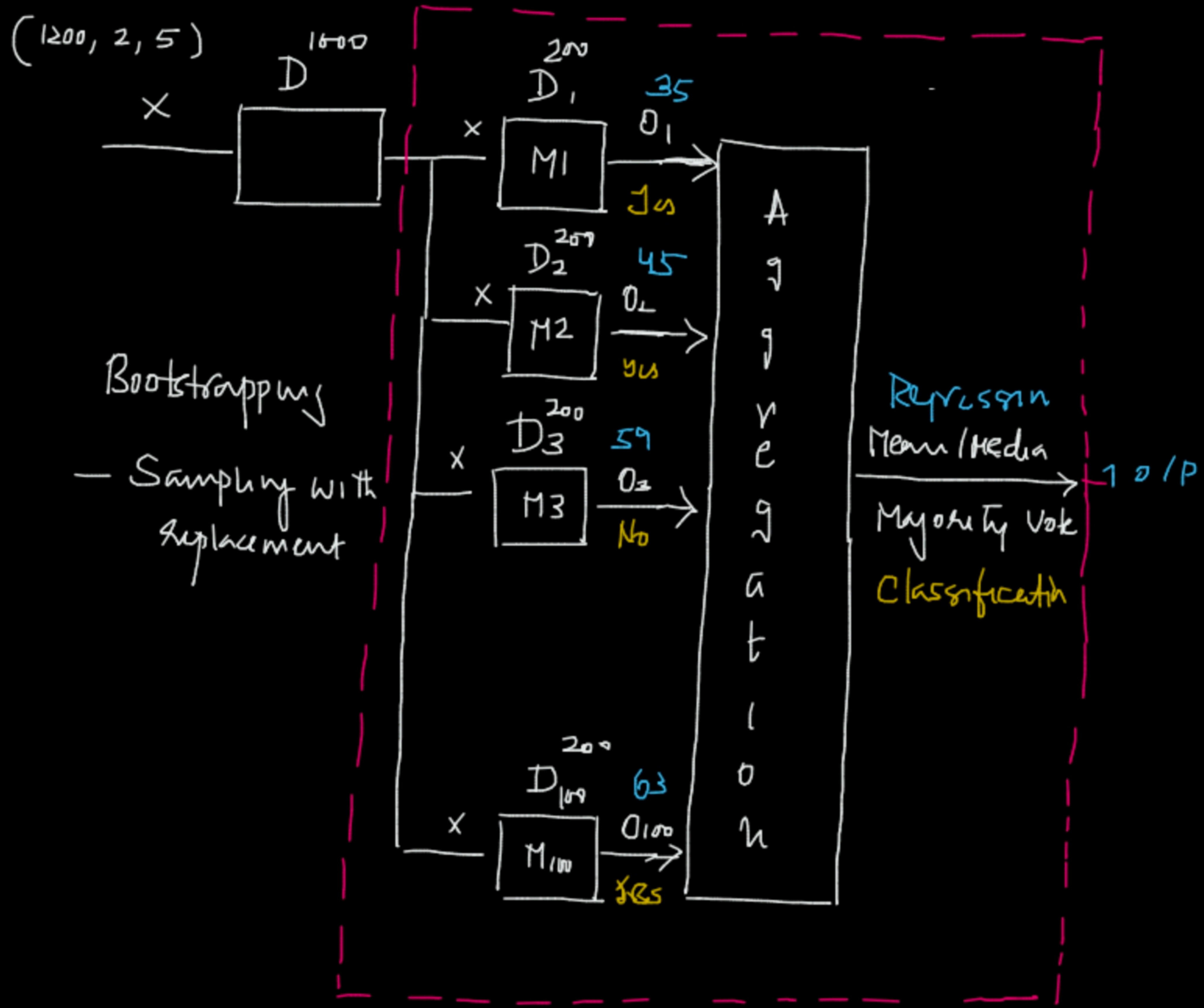


Ensemble Model

- collection of models
- Combining multiple models

- (a) Bagging ✓ → Technique — Random Forest
- (b) Boosting ✓ → Technique — Ada Boost, XGBoost
- (c) stacking
- (d) cascading

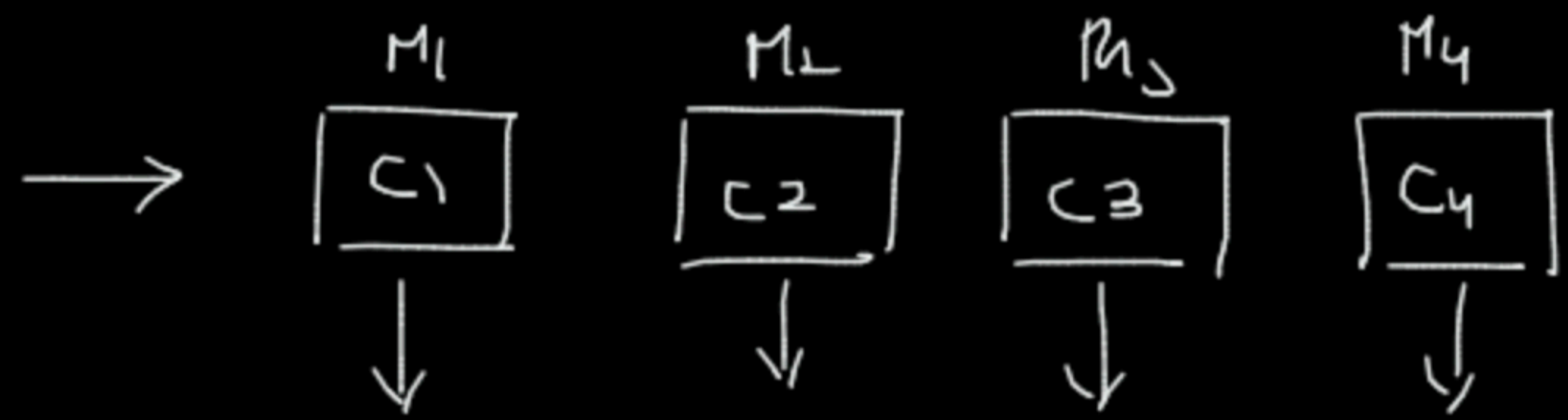
Bagging — Bootstrapped Aggregation



Random Forest

100 mins

- Models are indep of each other
Parallelisation is possible



→ 25 min

- Homogeneous Models
- Every model is given equal importance

✓ Bias & Variance →

10.37 am

← Data Encoding →

— Convert categorical data to Numeric data

```
le = LabelEncoder()
le.fit_transform(Xcol)
```

Label Encoding

Species

0

Setosa

1

Versicolva

2

Virginica

Alphabetical
order

oneHot Encoding

← dummy columns →
Species-Set Species-Vv Species-Virg

1

0 ✓

0 ✓

0 ✓

1

0 ✓

0 ✓

0 ✓

1

Sparse matrix

1 ohe = OneHotEncoder() →

ohe.fit_transform(X_→)

2 Pandas → get_dummies()

Stosn	1	0	0	✓
Var	0	1	0	✓
Vir	0	0	1	✓

0	0	0	-1
1	0	0	-2
0	1	0	-3
0	0	1	-4

$3 - 2\omega$
 $10 - 9\omega$