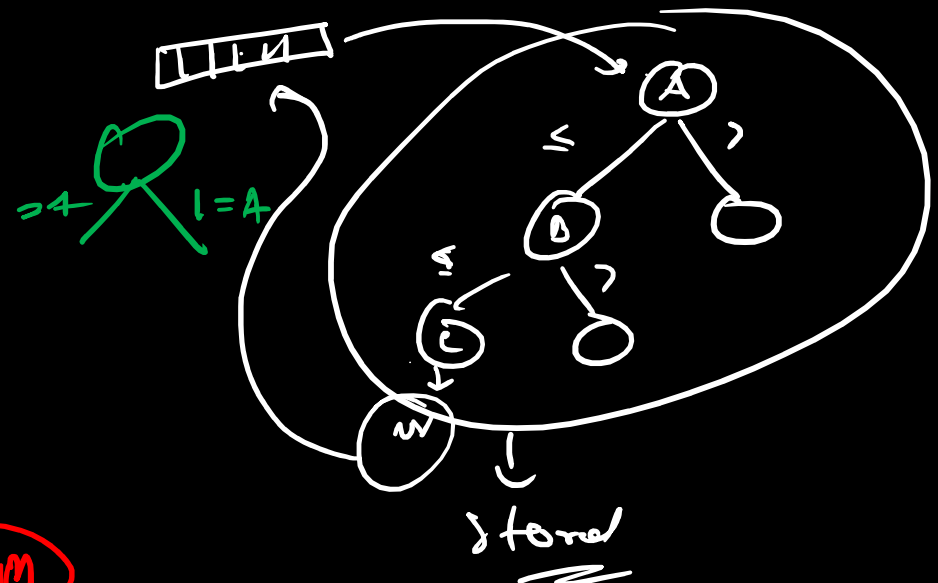
A
B
C
D

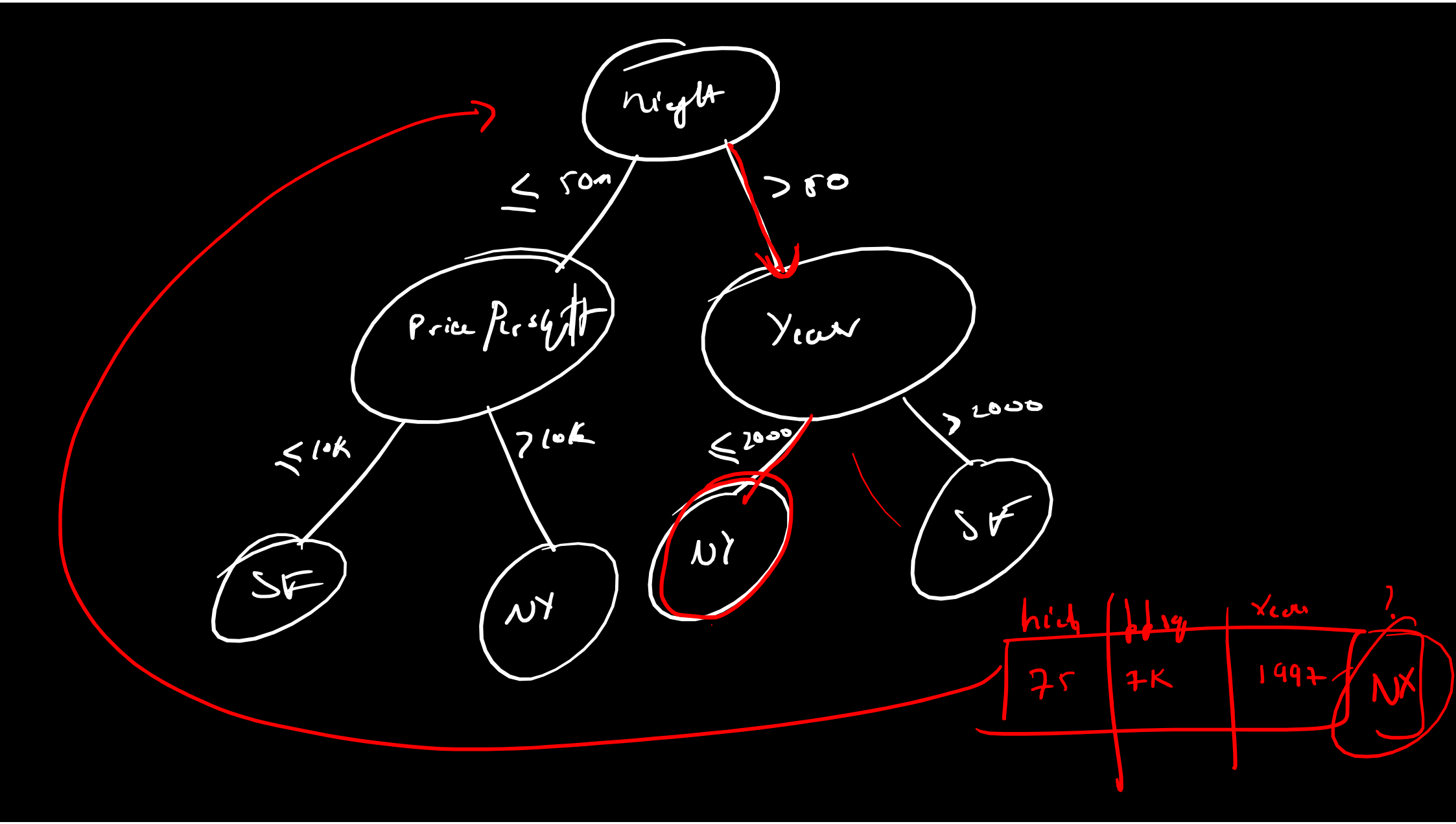
A



Data → Train → MLA → MLM

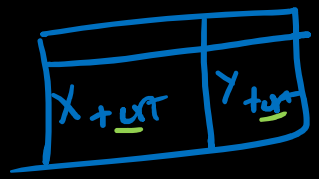
Data → Test → MLM → Prediction / Classification

492 → 250 → Decision Tree → Tree
↘ 242


Height
≤ 50m
> 50
Price Per Sqft
Year
≤ 10K
> 10K
≤ 2000
> 2000
SF
NY
NY
SF
height
Year
75
7K
1997
NY

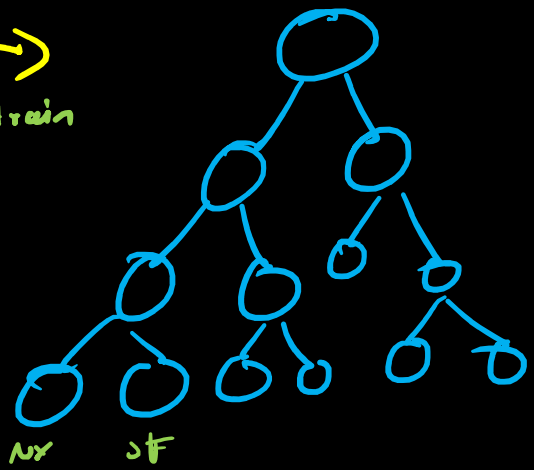
Training Data → 250 Rows

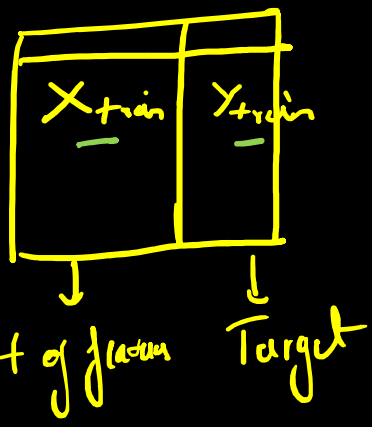
Testing Data → 242 Rows

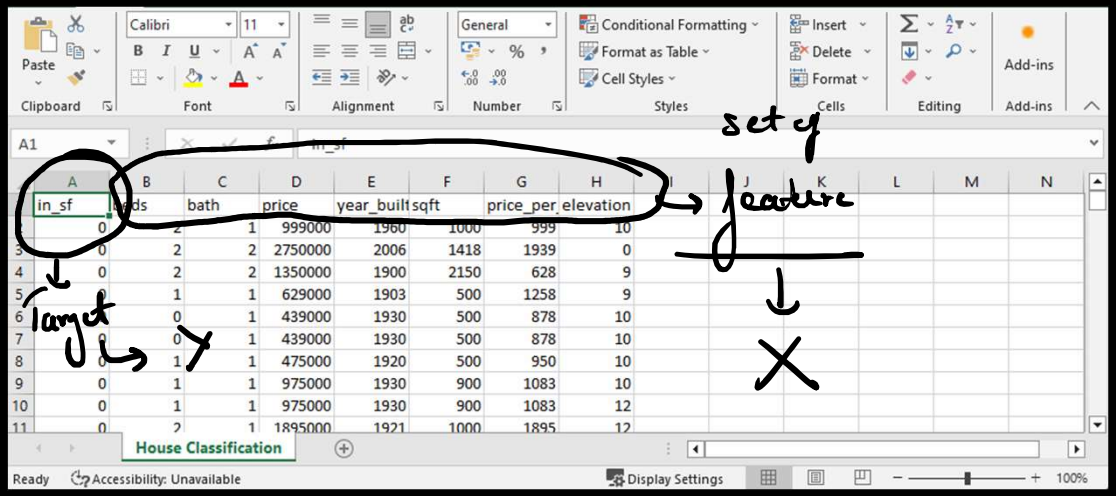


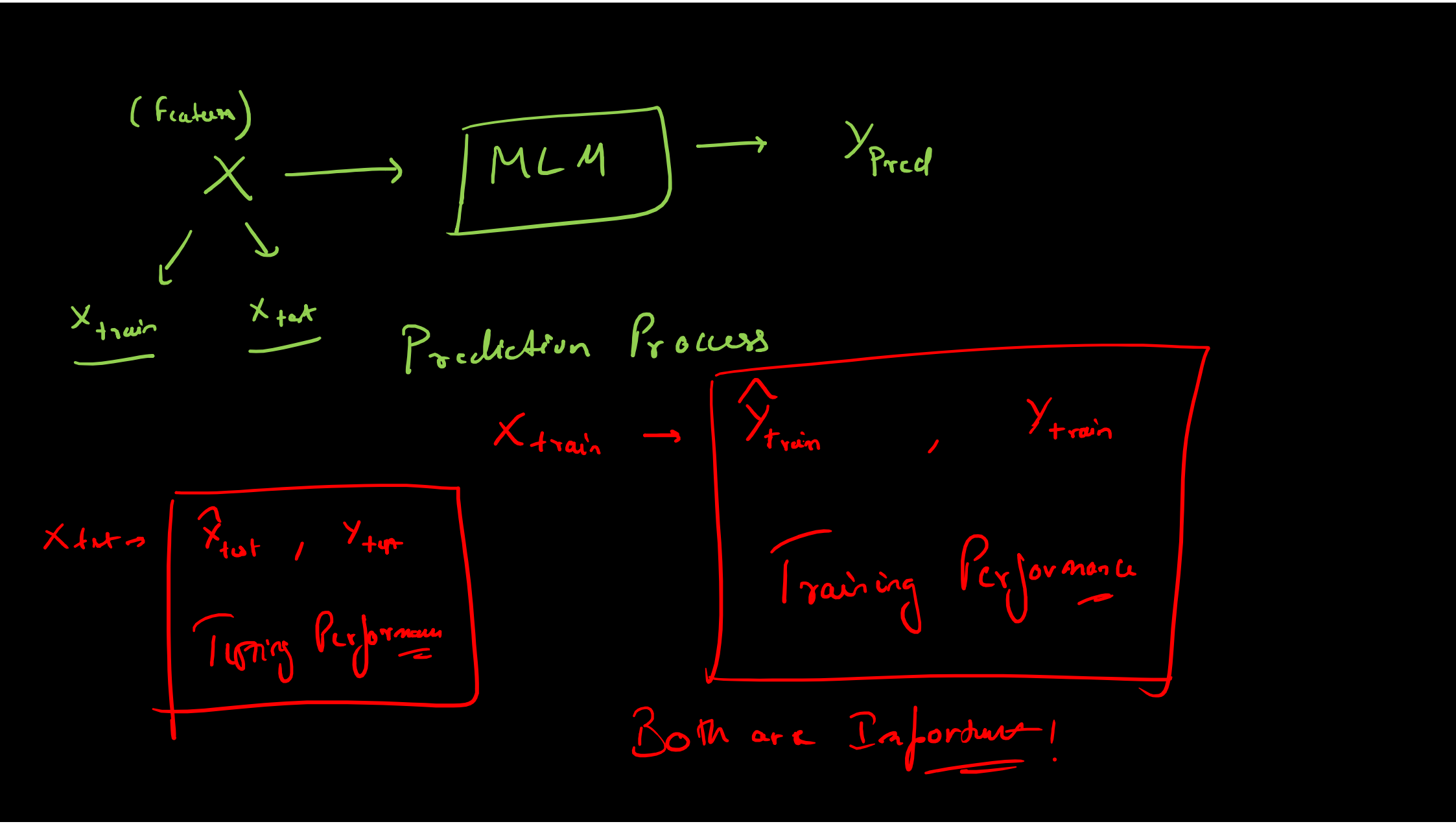
Training

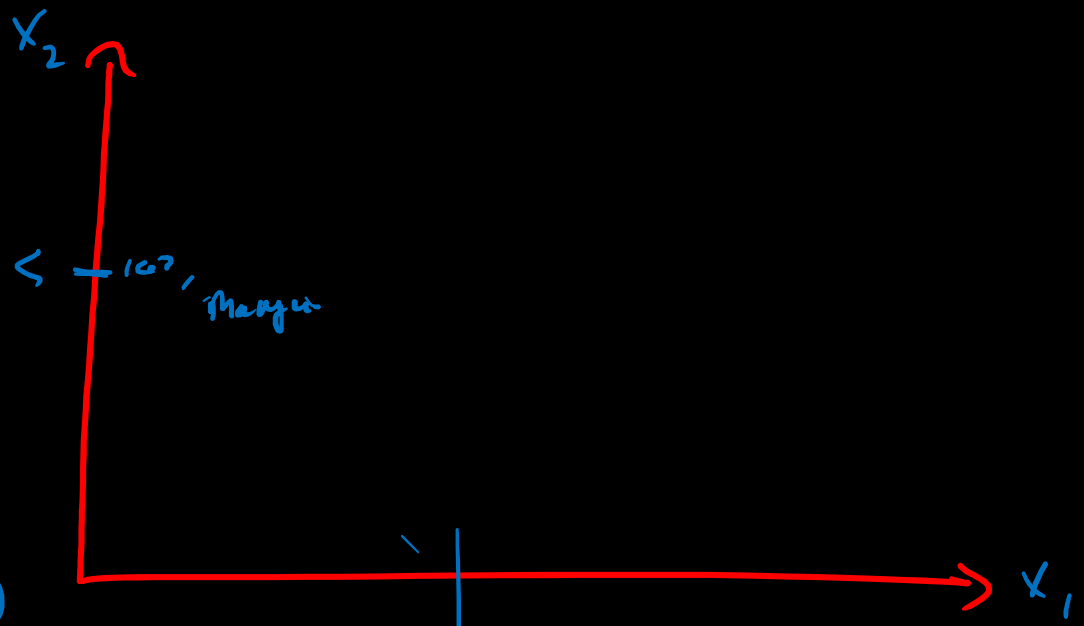
X_{train} , Y_{train}








(Features)
X
MLM
Y_{Pred}
X_{train}
X_{test}
Prediction Process
X_{train}
$\hat{Y}_{train}$, Y_{train}
Training Performance
X_{test}
$\hat{Y}_{test}$, Y_{test}
Testing Performance
Both are Important!



X_1 → Age (18-50)

500 → 18-50

~~36~~ 300 → 18-50

200 → 18-50

34 → Engg → Need for X

Strict
↳ bounding

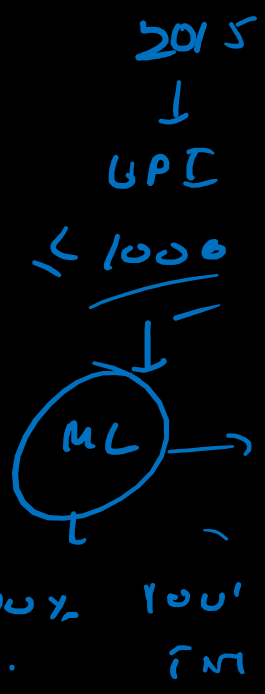
long-term predictions ?

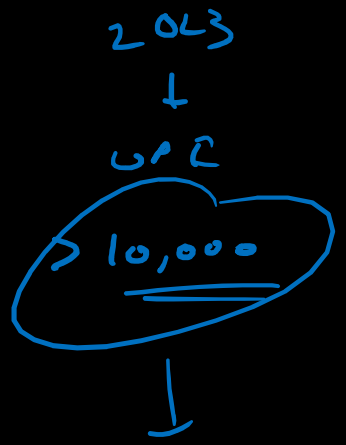
Training Acc is 100%.

↓

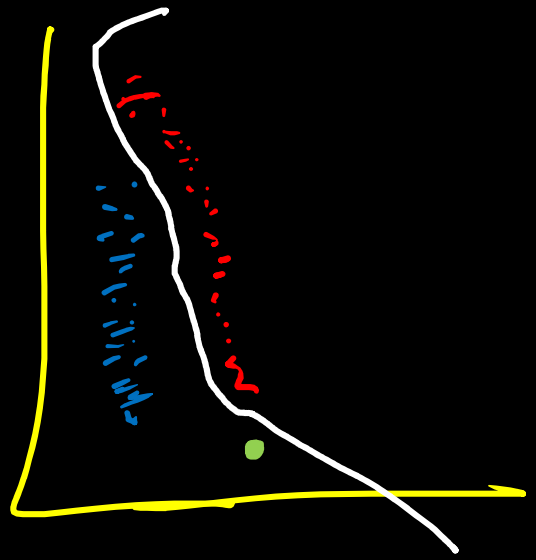
Strict boundary?

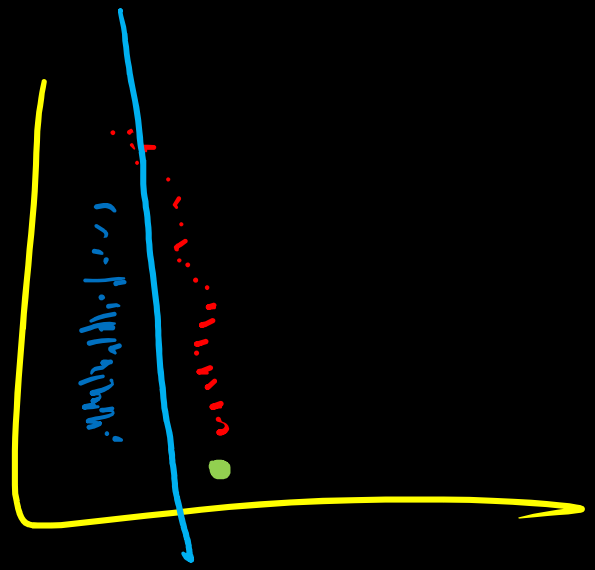
Yes





≥ 95% → strict boundry
↳ Not good for long-term predictions





Overfitting → Training Acc ≥ 97%

Training performance → Imp?? → To identify how good the model has learnt.

≥ 95% → overfitting

≤ 50% → underfitting

Use-case Threshold → compared with Testing Acc ≥ Threshold

↓

Deployed/not