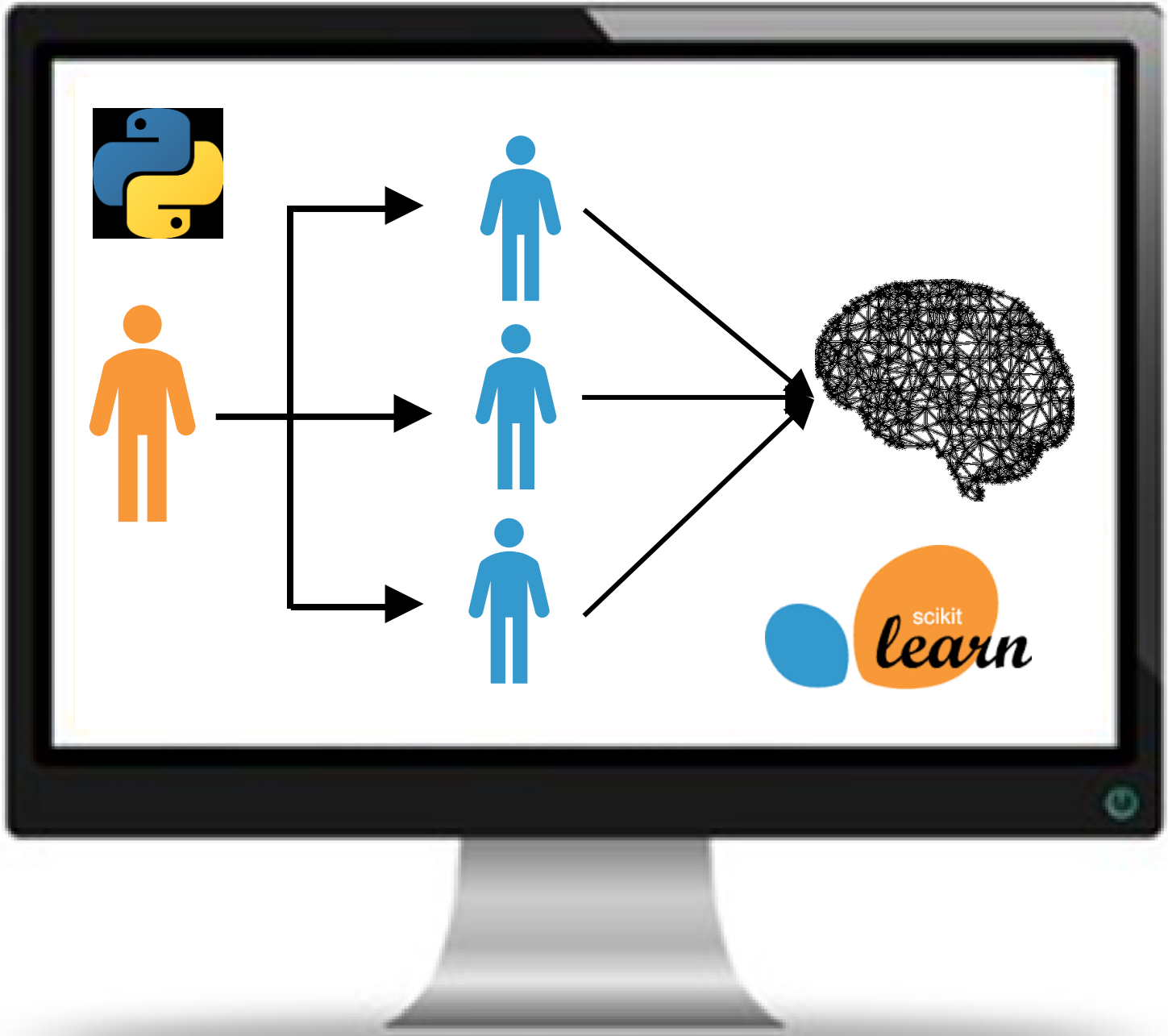

scikit
learn

ENSEMBLE MACHINE LEARNING

By Ankit Mistry

COURSE OVERVIEW

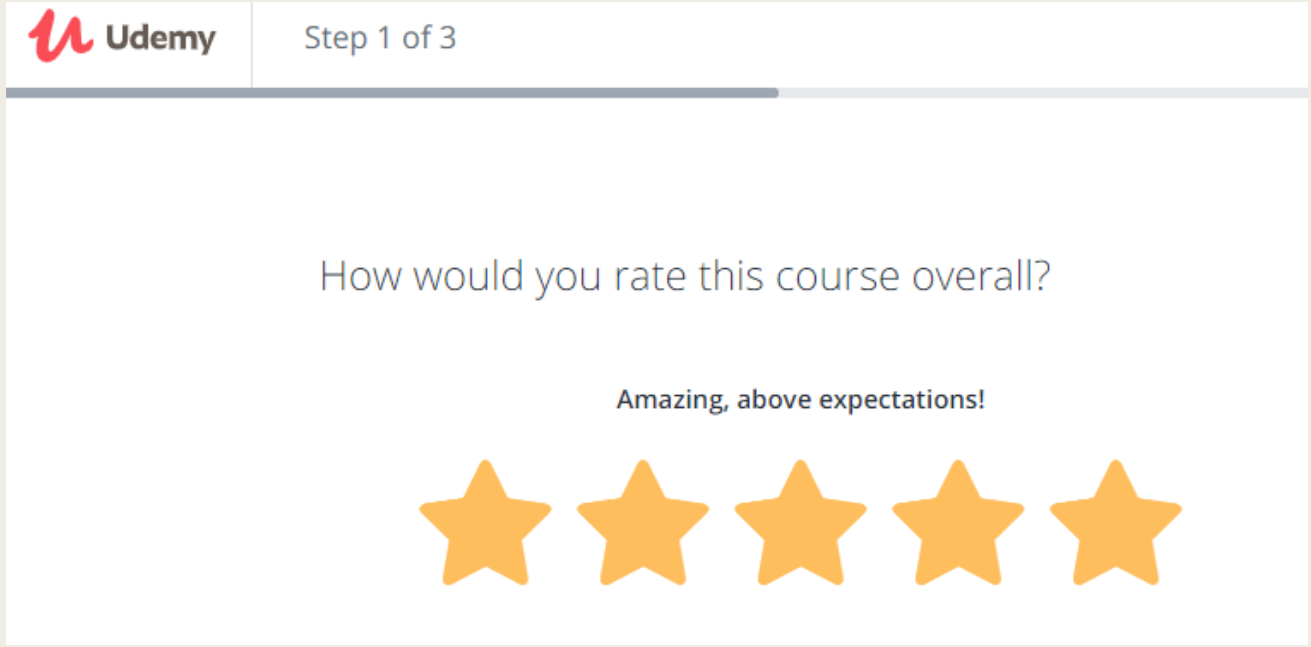
By Ankit Mistry

Prerequisites

- Basics of Python Programming
- Machine Learning Algorithms
- Build model using Scikit-learn library

FAQ

- Download Video Offline
- Video Streaming Quality
  - *Suggest to watch in 720P Or Higher*
- Playback Speed
- Reviews
- support@udemy.com
- Q & A Forum



SECTION 1

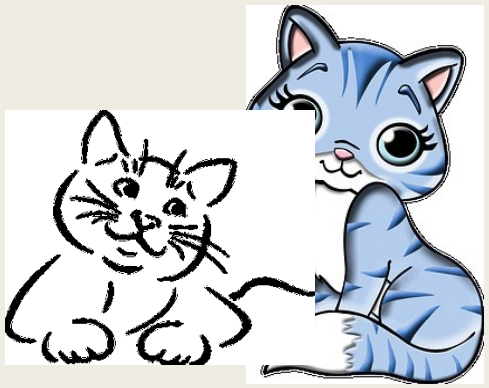
SECTION INTRODUCTION

- Machine Learning
- Bias and Variance Tradeoff
- Ensemble Learning
- Type of Ensemble Learning Methods

MACHINE LEARNING

By Ankit Mistry

Machine Learning



Machine Learning Algorithm

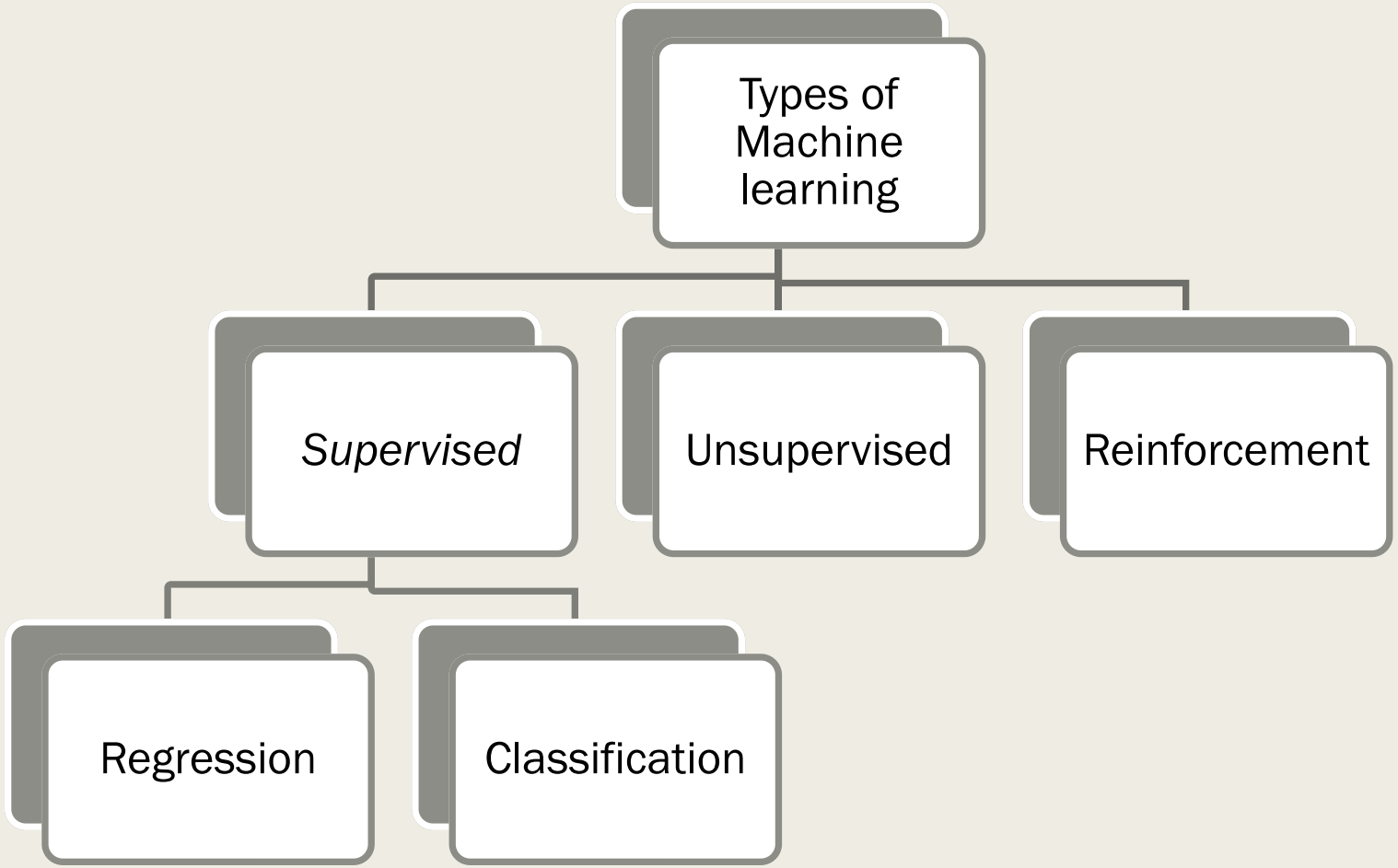
Model

Cat

By Ankit Mistry

https://cdn.pixabay.com/photo/2014/11/30/14/11/kitty-551554__340.jpg

Types of Machine learning



By Ankit Mistry

Machine Learning Workflow

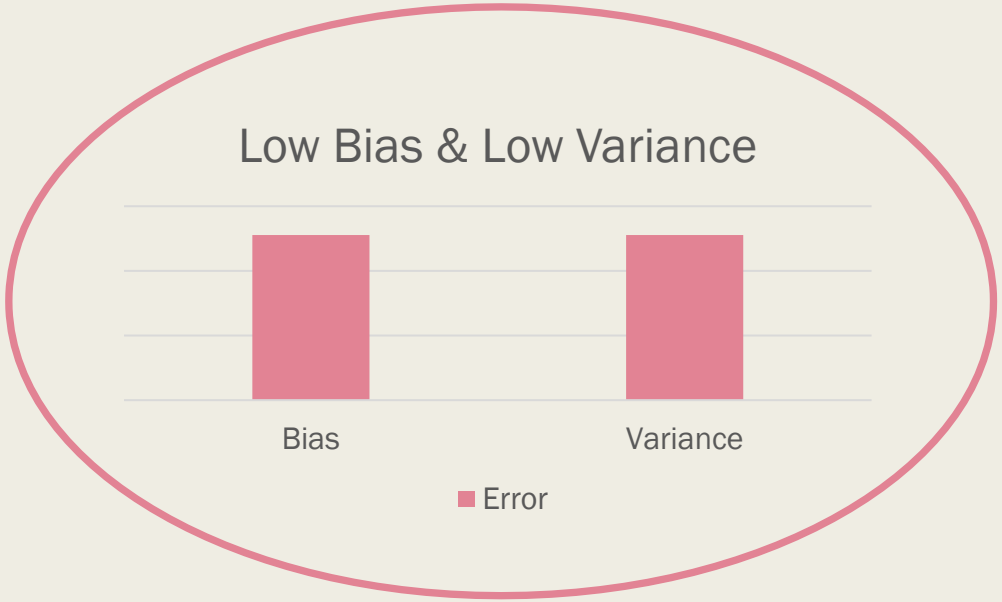
By Ankit Mistry

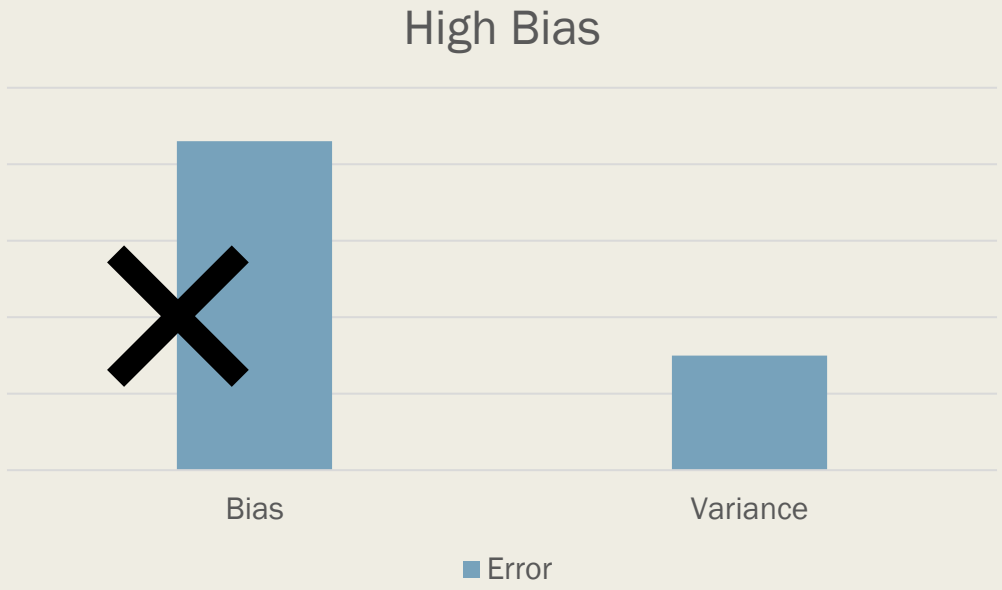
BIAS VARIANCE TRADE OFF

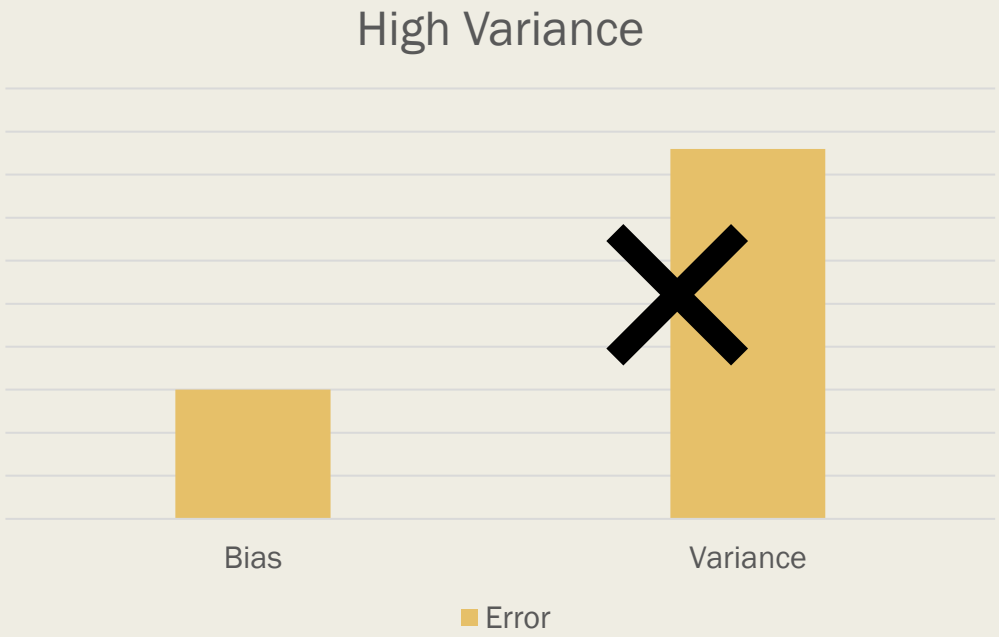
By Ankit Mistry

Bias & Variance

- Bias Error
- Variance Error
- Error which can not be reduced





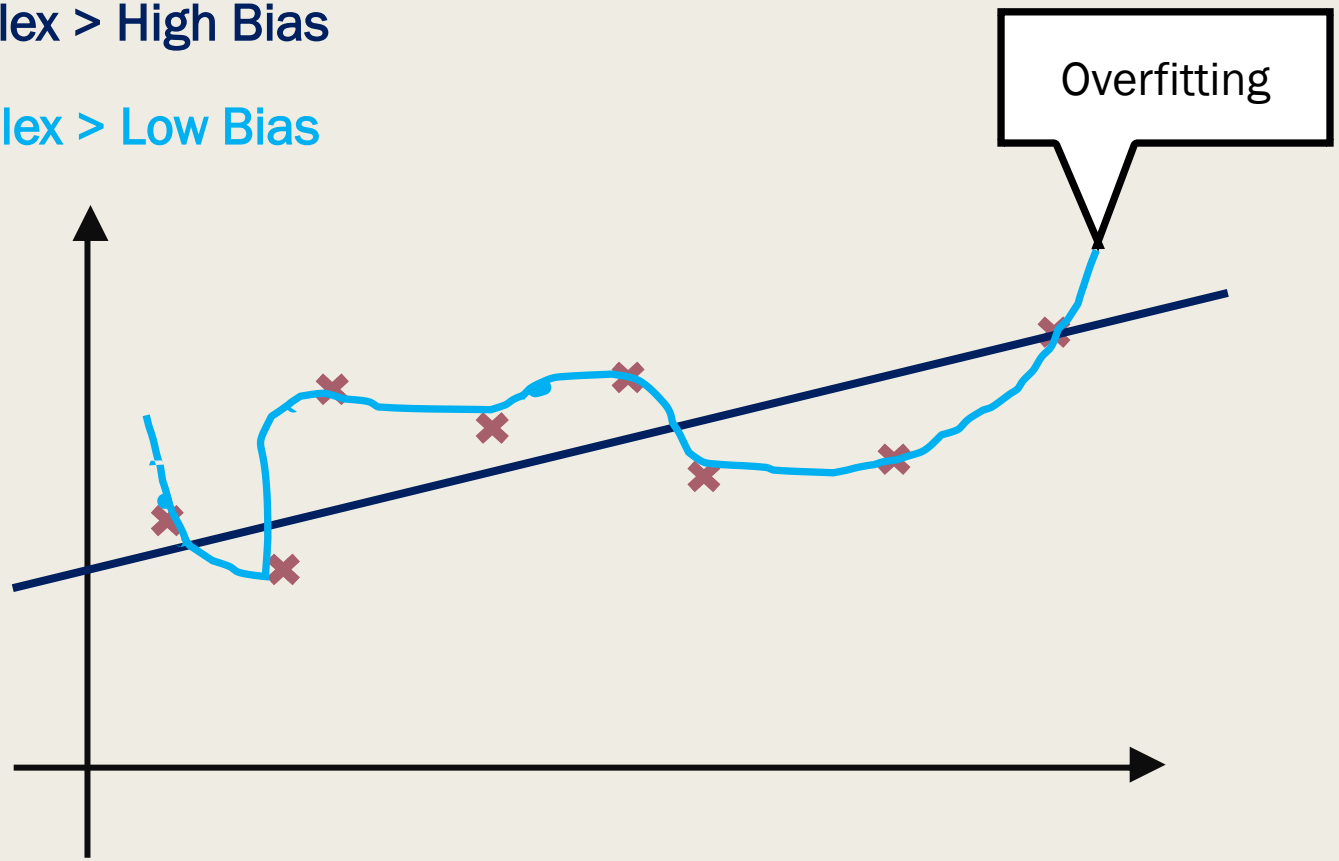


By Ankit Mistry

Bias

- Assumptions about model
- **More assumption > Less Complex > High Bias**
- **Less assumption > More Complex > Low Bias**

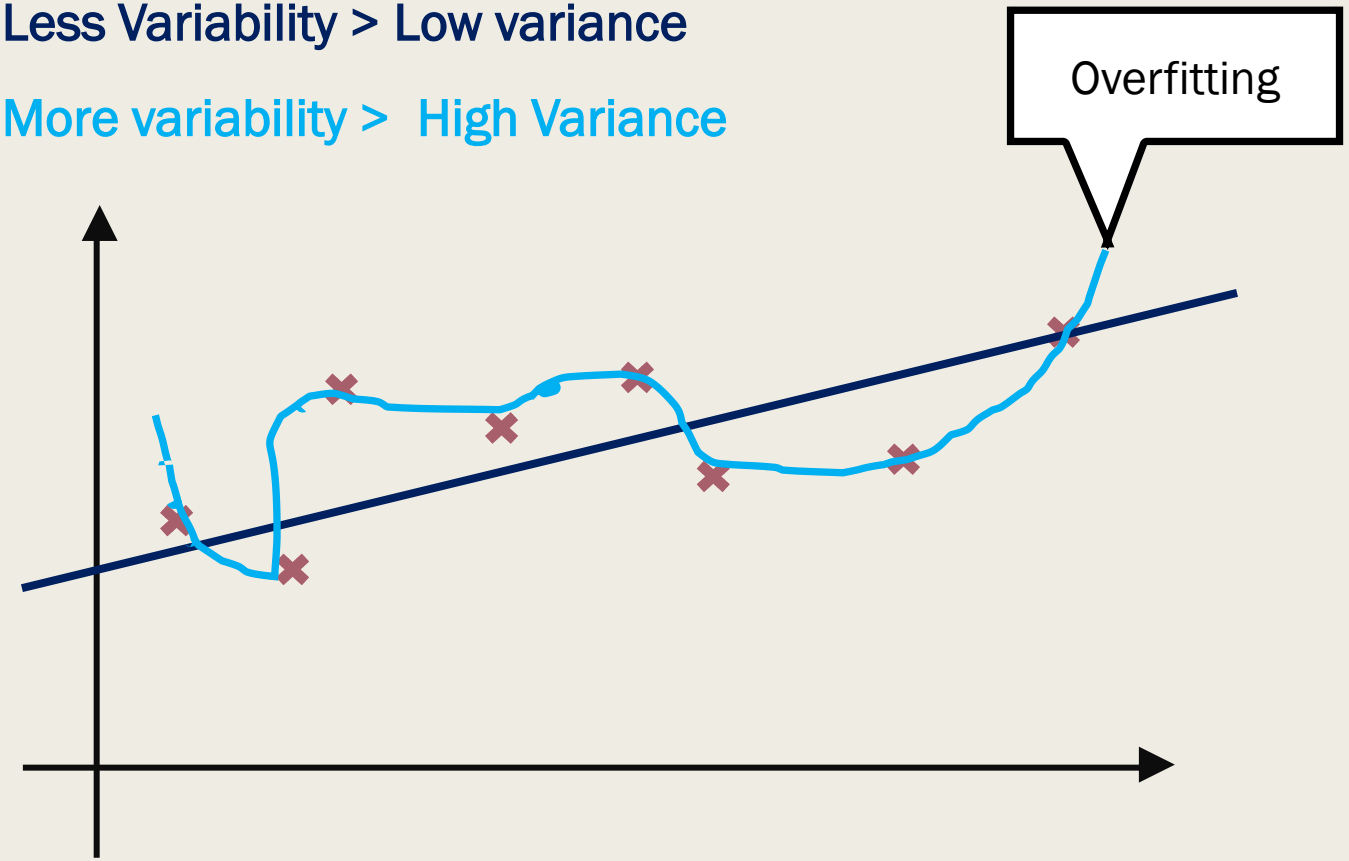
- Linear Regression
- Linear Discriminant Analysis
- Logistic Regression



Variance

- Model Fluctuation on Data Change
- **More assumption > Less Complex > Less Variability > Low variance**
- **Less assumption > More Complex > More variability > High Variance**

- Decision Trees
- K-Nearest Neighbors
- Support Vector Machines.



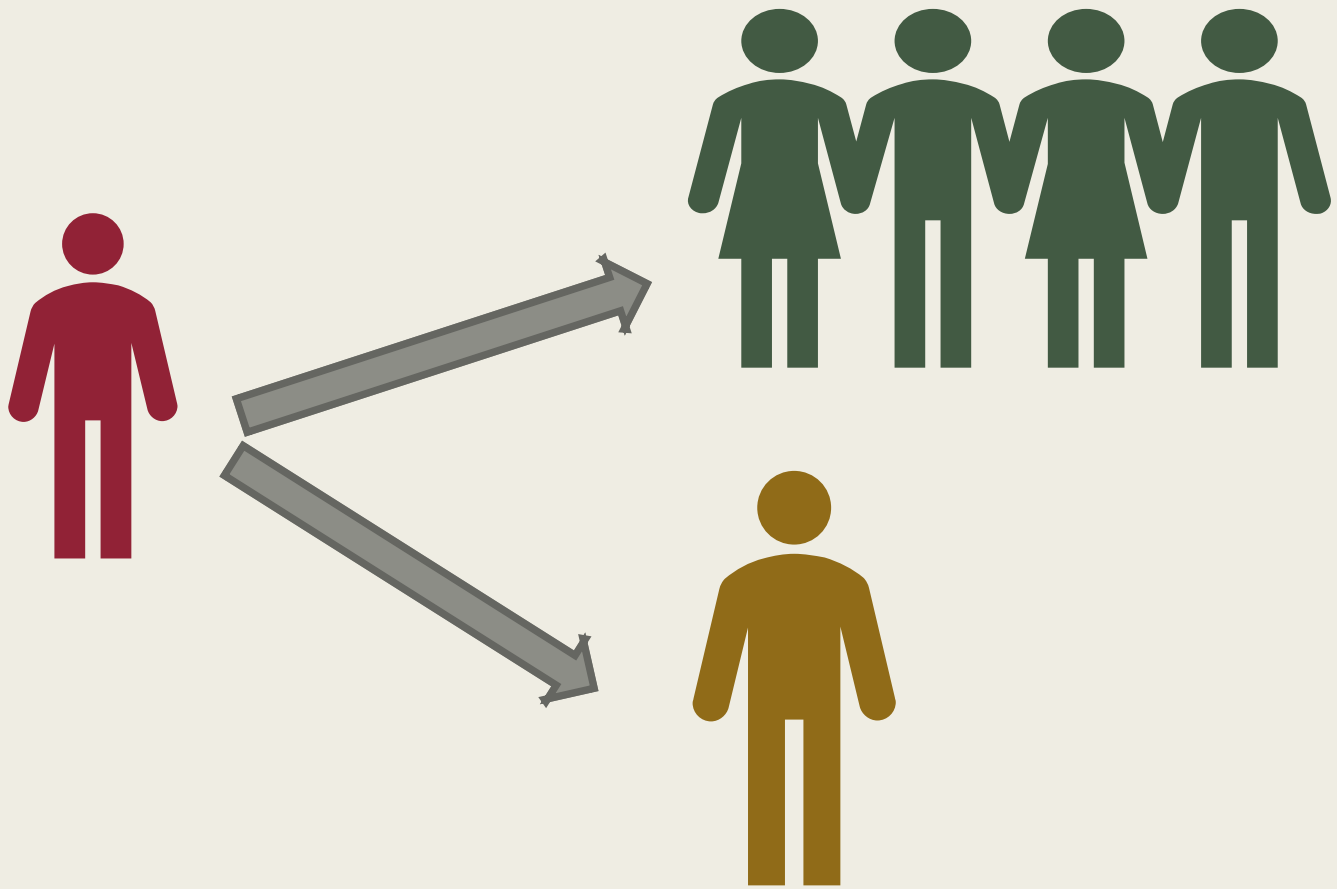
HOW TO REDUCE BIAS – VARIANCE ERROR

By Ankit Mistry

ENSEMBLE LEARNING

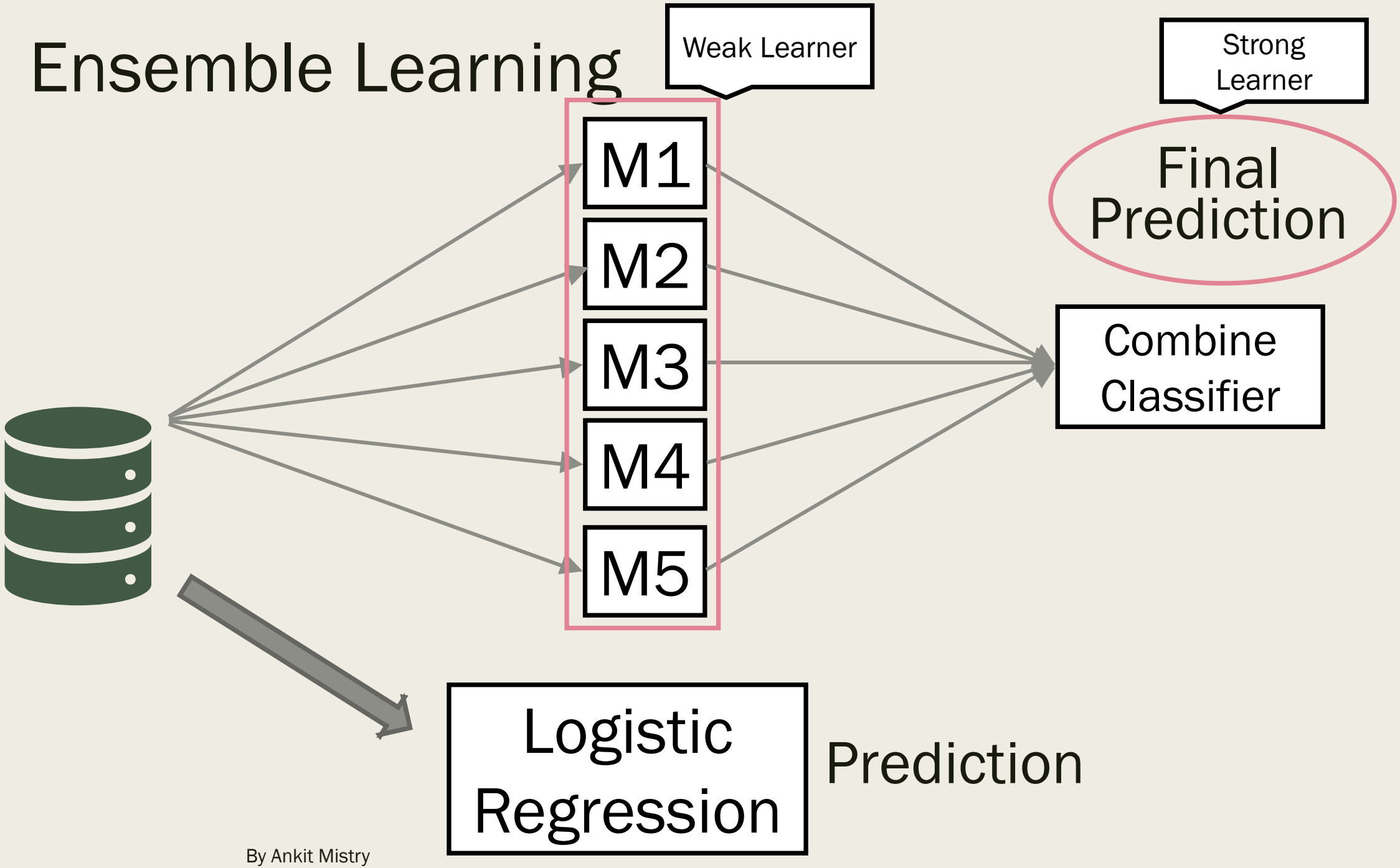
By Ankit Mistry

Ensemble Learning



By Ankit Mistry

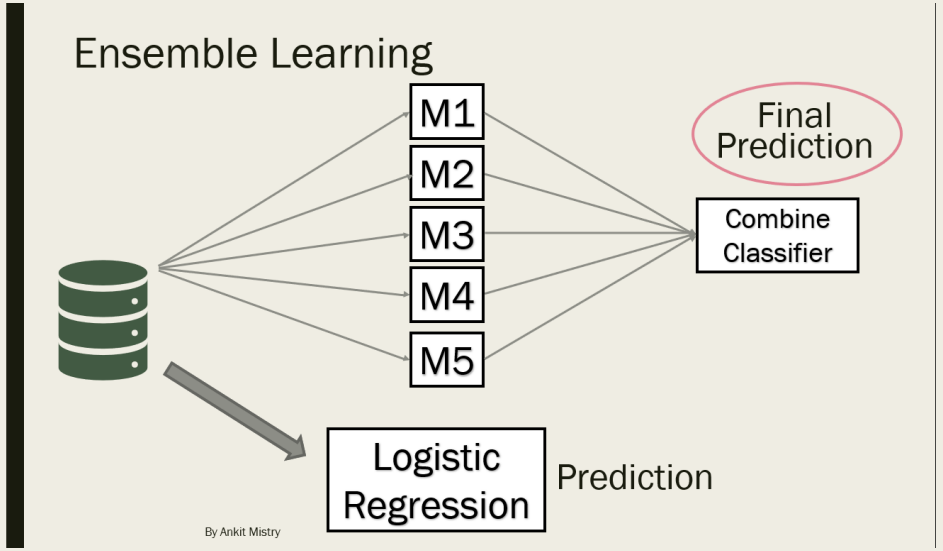
Ensemble Learning



By Ankit Mistry

Ensemble Learning Advantage

- Turn weak classifier into strong
- Single model is biased
- Variance error reduces
- Model accuracy increases
- Less overfitting

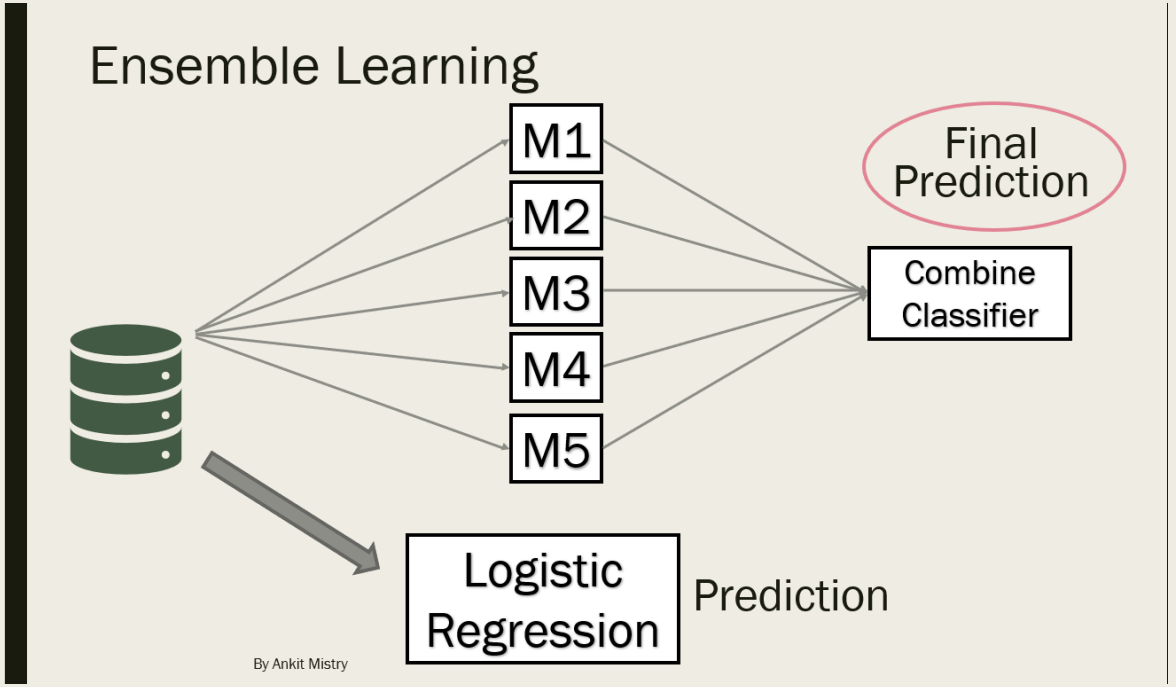


Disadvantage

- High computational cost

Questions

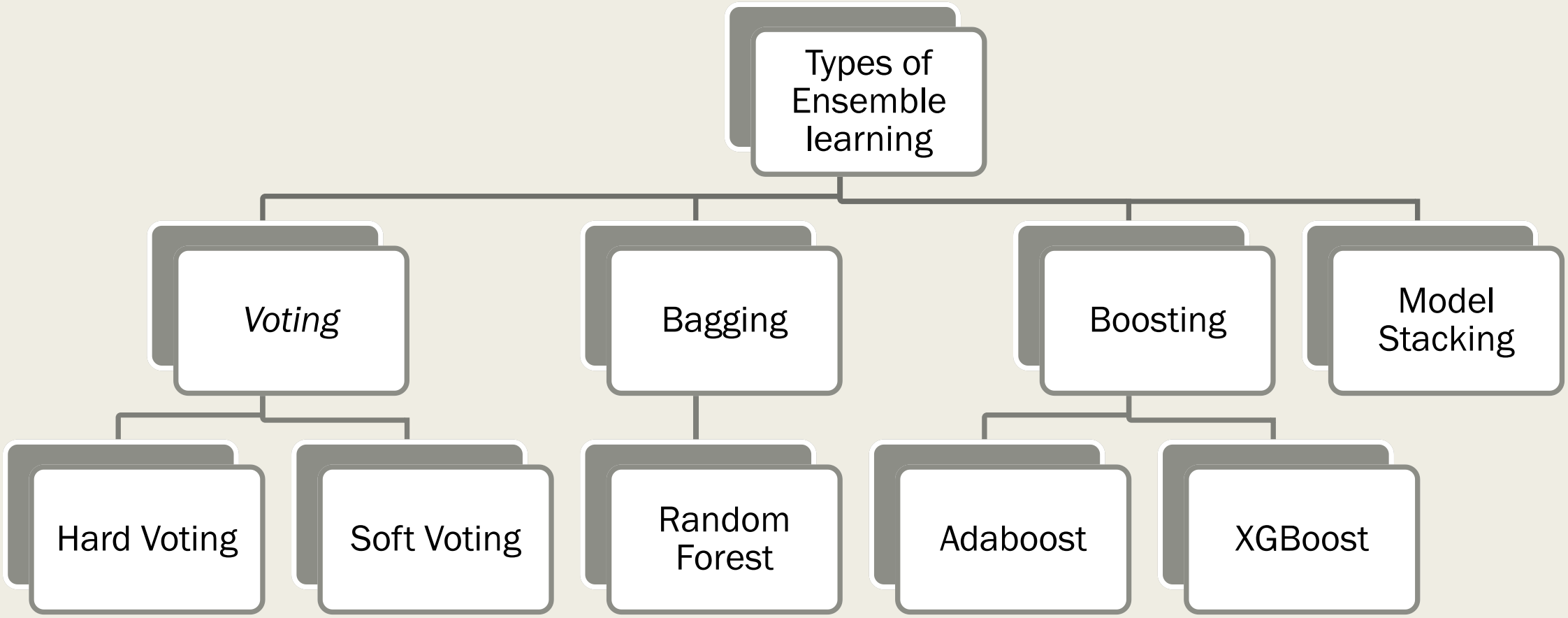
- Do we use all training data on every model
- What is M1, M2, M3 …
- How to train all model
- How to combine output from all model



TYPES OF ENSEMBLE LEARNING

By Ankit Mistry

Types of Ensemble learning



By Ankit Mistry

SECTION 2

Section 2

- You are free to use any IDE of your choice.
- Install Anaconda
- Install Library
  - *Scikit-learn*
  - *Jupyter Notebook*

DOWNLOAD & INSTALL ANACONDA

By Ankit Mistry

INSTALL LIBRARY SCIKIT-LEARN JUPYTER NOTEBOOK

By Ankit Mistry

Install Library

- Package manager
  - *pip install pkg_name* *pip install scikit-learn==0.22.1*
  - *Conda install pkg_name*

REGRESSION PROBLEM

SCIKIT-LEARN WORKFLOW

By Ankit Mistry

CLASSIFICATION PROBLEM

SCIKIT-LEARN WORKFLOW

By Ankit Mistry

SECTION 3

Section 3

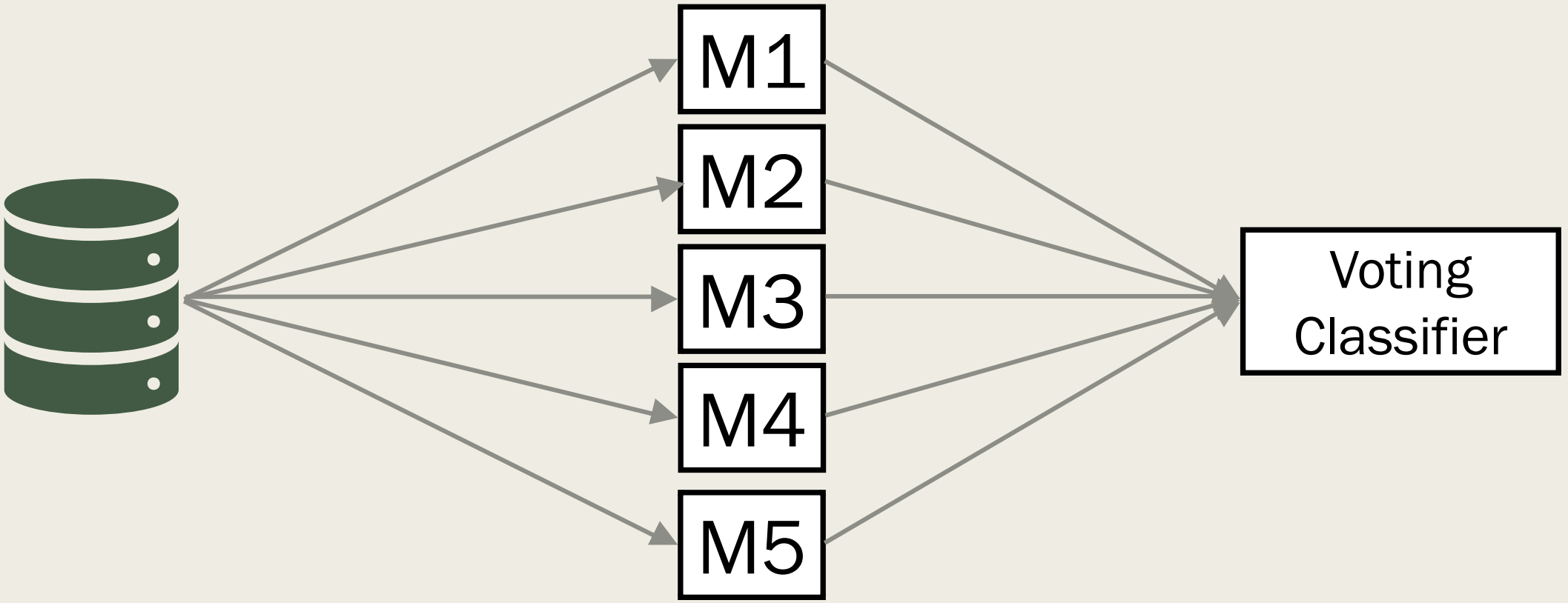
- Voting classifier
  - *Hard*
  - *Soft*
- Averaging and Weighted Averaging
- [Hands-on] Voting classifier

VOTING CLASSIFIER

By Ankit Mistry

Voting Classifier

- Do we use all training data on every model
- What is M1, M2, M3 …
- How to train all model
- How to combine output from all model



Voting Classifier

- Do we use all training data on every model
  - *Yes*
- What is M1, M2, M3 …
  - *As different as Possible :*
- How to train all model
  - *Train in Parallel*
- How to combine output from all model with voting classifier

Voting Classifier

- **Classification**
  - *Hard Voting*
  - *Soft Voting*
- Regression
  - *Averaging*

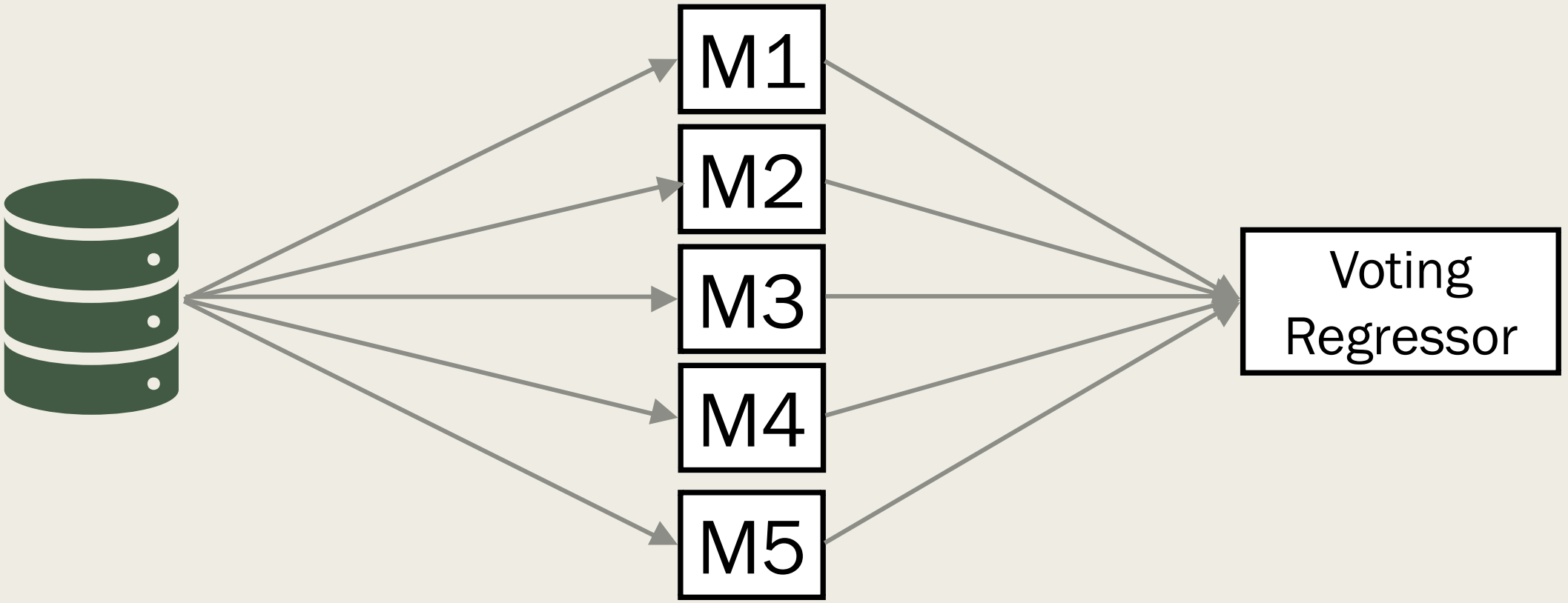
	Hard Voting	Soft Voting	
M1	1	0.1, 0.9	0.2, 0.8
M2	0	0.95, 0.05	0.6, 0.4
M3	1	0.3, 0.7	0.75, 0.25
M4	1	0.2, 0.8	0.4, 0.6
M5	1	0.45, 0.55	0.62, 0.38
	1	0.4, 0.6	0.514, 0.486
		1	0

VOTING REGRESSOR

By Ankit Mistry

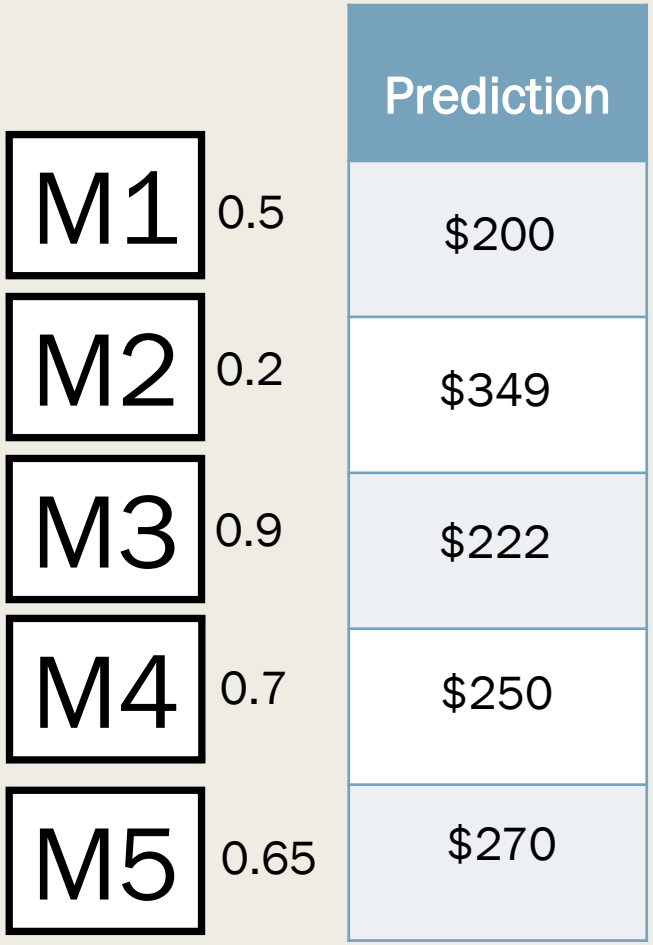
Voting Regressor

- Do we use all training data on every model
- What is M1, M2, M3 …
- How to train all model
- How to combine output from all model



Averaging

- Classification
  - *Hard Voting*
  - *Soft Voting*
- **Regression**
  - *Averaging*



$$\text{Average} = \frac{200 + 349 + 222 + 250 + 270}{5}$$

$$= 258.2$$

$$\text{Weighted Average} = \frac{0.5*200 + 0.2*349 + 0.9*222 + 0.7*250 + 0.65*270}{5}$$

$$= 144.02$$

VOTING CLASSIFIER WITH SCIKIT-LEARN

By Ankit Mistry

Voting with Scikit-learn

- Apply 3 Classification algorithm of credit card data
- Apply ensemble.**VotingClassifier** with **soft** Voting
- Apply ensemble.**VotingClassifier** with **Hard** Voting

- **VotingRegressor** as Exercise

VOTING REGRESSOR WITH SCIKIT-LEARN

By Ankit Mistry

SECTION 4

Section 4

- Concept of Bagging Ensemble
- [Hands-on] Bagging classifier
- Random Forest
- [Hands-on] Random Forest

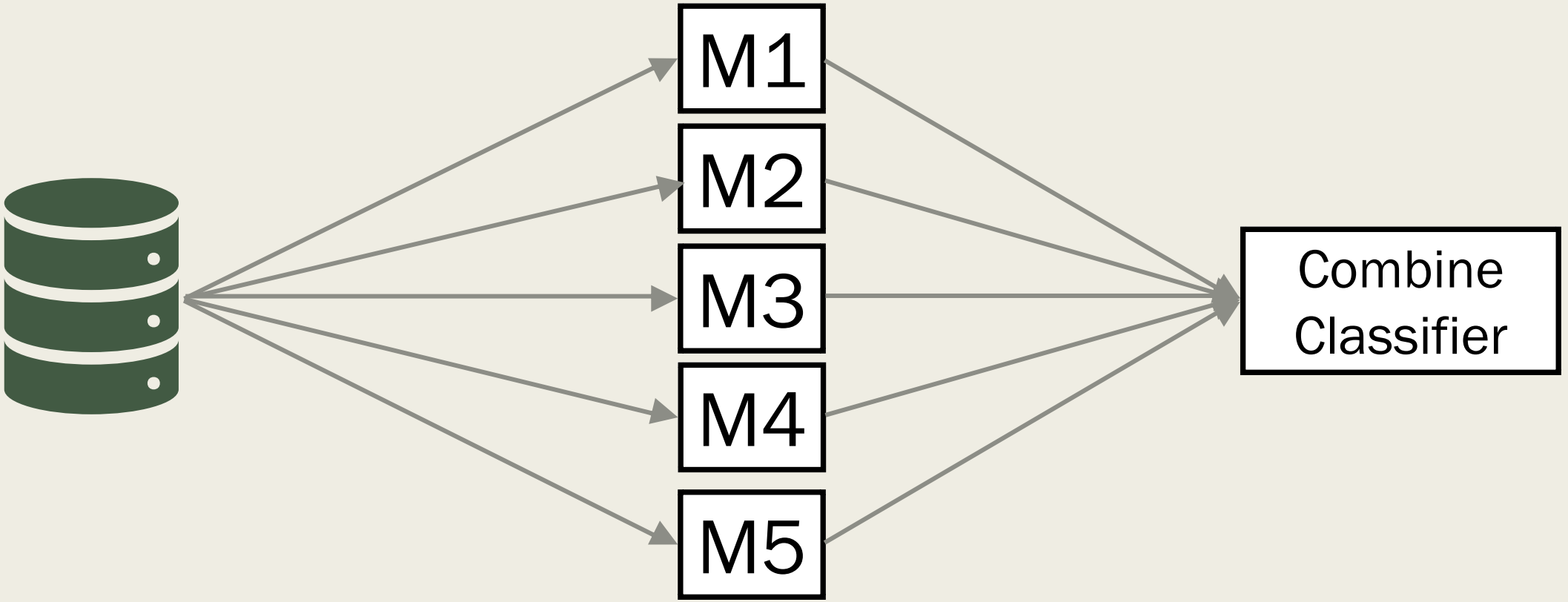
BAGGING

Bootstrap Aggregation

By Ankit Mistry

Bagging

- Do we use all training data on every model
- What is M1, M2, M3 …
- How to train all model
- How to combine output from all model



Bagging Ensemble

- Do we use all training data on every model
  - *No*
  - *Then How to select data*
- What is M1, M2, M3 …
  - *Same weak learner ML Algorithm*
- How to train all model
  - *Train in Parallel*
- How to combine output from all model
  - *voting classifier*

How to Select Data For Training

- Row Sampling
- Feature Sampling
- Combine above both (Row + Feature)

How to select Data Row Sampling

	C1	C2	C3	C4
R0				
R1				
R2				
R3				
R4				
R5				
R6				
R7				
R8				
R9				

Random Row Sample with Replacement

	C1	C2	C3	C4
R1				
R7				
R8				
R2				
R0				

M1

	C1	C2	C3	C4
R4				
R1				
R2				
R6				
R5				

M2

	C1	C2	C3	C4
R1				
R0				
R3				
R7				
R5				

M3

By Ankit Mistry

How to select Data Feature Sampling

	C1	C2	C3	C4
R0				
R1				
R2				
R3				
R4				
R5				
R6				
R7				
R8				
R9				

Random Feature Sample with Replacement

	C1	C2
R0		
R1		
R2		
R3		
R4		
R5		
R6		
R7		
R8		
R9		

M1

	C1	C2	C3
R0			
R1			
R2			
R3			
R4			
R5			
R6			
R7			
R8			
R9			

M2

	C3	C4
R0		
R1		
R2		
R3		
R4		
R5		
R6		
R7		
R8		
R9		

M3

By Ankit Mistry

Row + Feature Sampling

	C1	C2	C3	C4
R0				
R1				
R2				
R3				
R4				
R5				
R6				
R7				
R8				
R9				

Random Feature Sample with Replacement

	C3	C4
R2		
R3		
R4		
R5		

M1

	C1	C2	C3	C4
R7				
R8				
R9				

M2

	C3	C4
R0		
R1		
R2		
R3		
R4		
R5		
R6		
R7		
R8		
R9		

M3

By Ankit Mistry

What is weak Leaner (M1, M2, ...)

- Classification
  - *SVM*
  - *Decision Tree*
  - *Logistic Regression*
  - *KNN*
  - *Neural Network*
- Regression
  - *Linear or Multiple Regression*
  - *Lasso – Ridge Regression*
  - *Polynomial Regression*

Random Forest

- Random Forest is special version of Bagging ensemble.
- Its is ensemble of Decision Tree

Bootstrap aggregation

- Do we use all training data on every model
  - Bootstrapping Sample >**Bootstrap aggregation**

[HANDS-ON] BAGGING ENSEMBLE

By Ankit Mistry

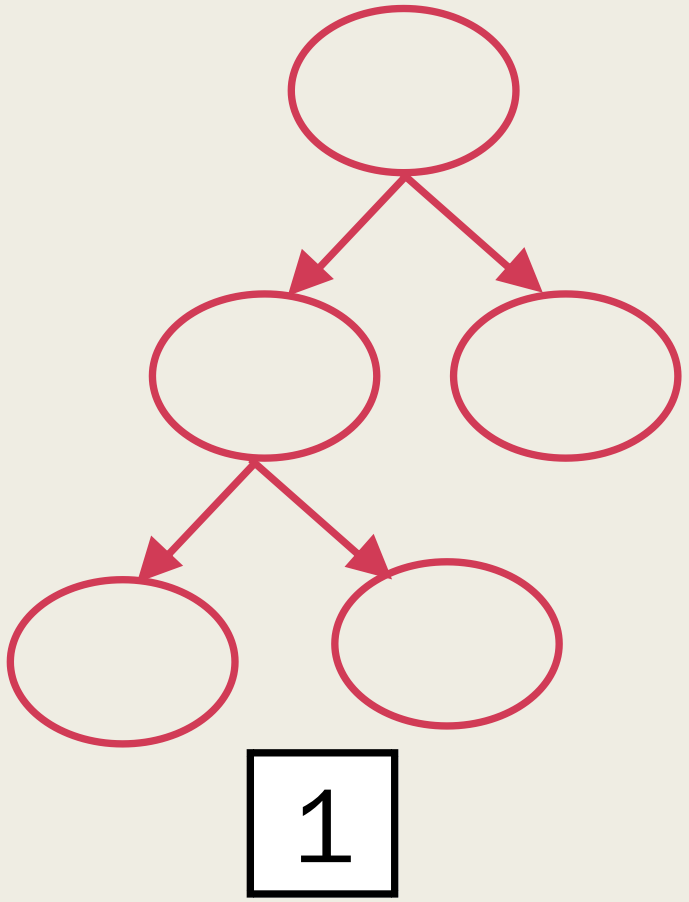
RANDOM FOREST

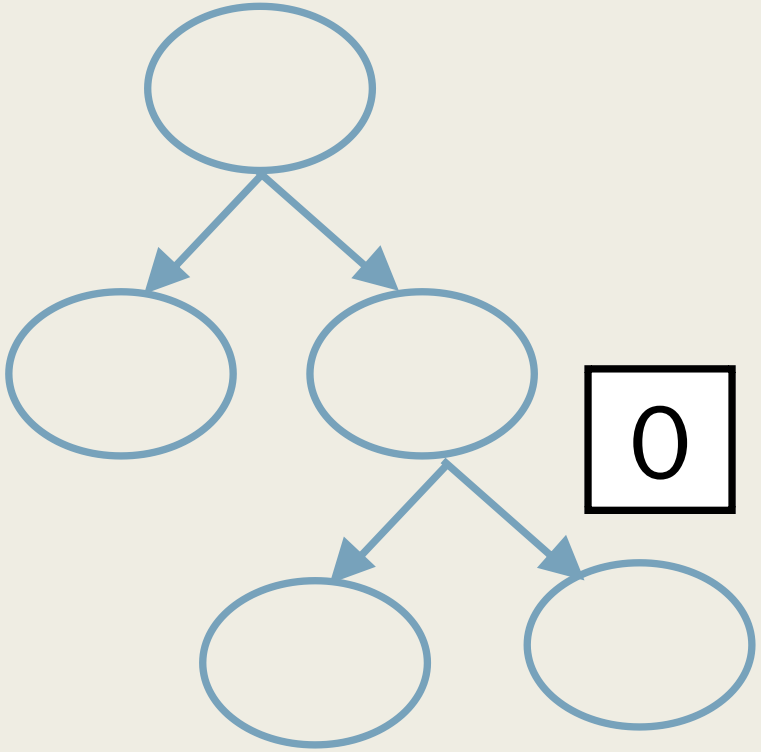
By Ankit Mistry

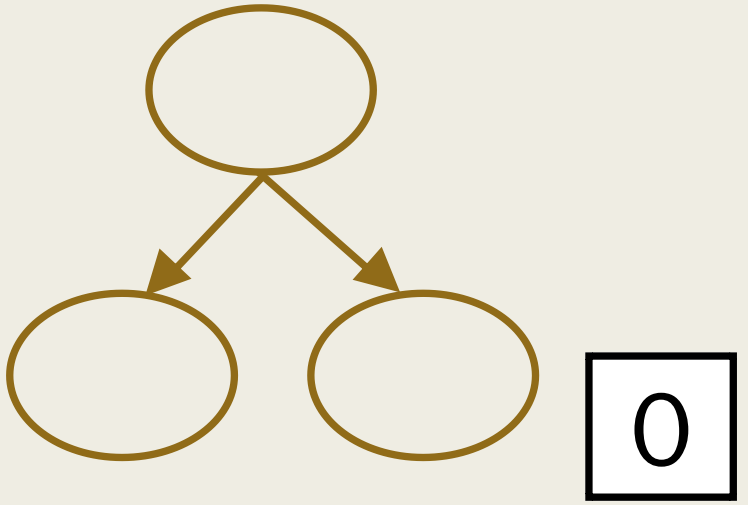
Random Forest

- Random Forest is ensembles of Decision Tree

	C1	C2	C3	C4
R0				
R1				
R2				
R3				
R4				
R5				
R6				
R7				
R8				
R9				







By Ankit Mistry

[HANDS-ON] RANDOM FOREST

By Ankit Mistry

SECTION 5

By Ankit Mistry

Section 5

- Concept of Boosting

- Adaboost (Adaptive Boosting)
- [Hands-on] Adaboost

- Gradient Boosting Machine (GBM)
- [Hands-on] Gradient Boosting

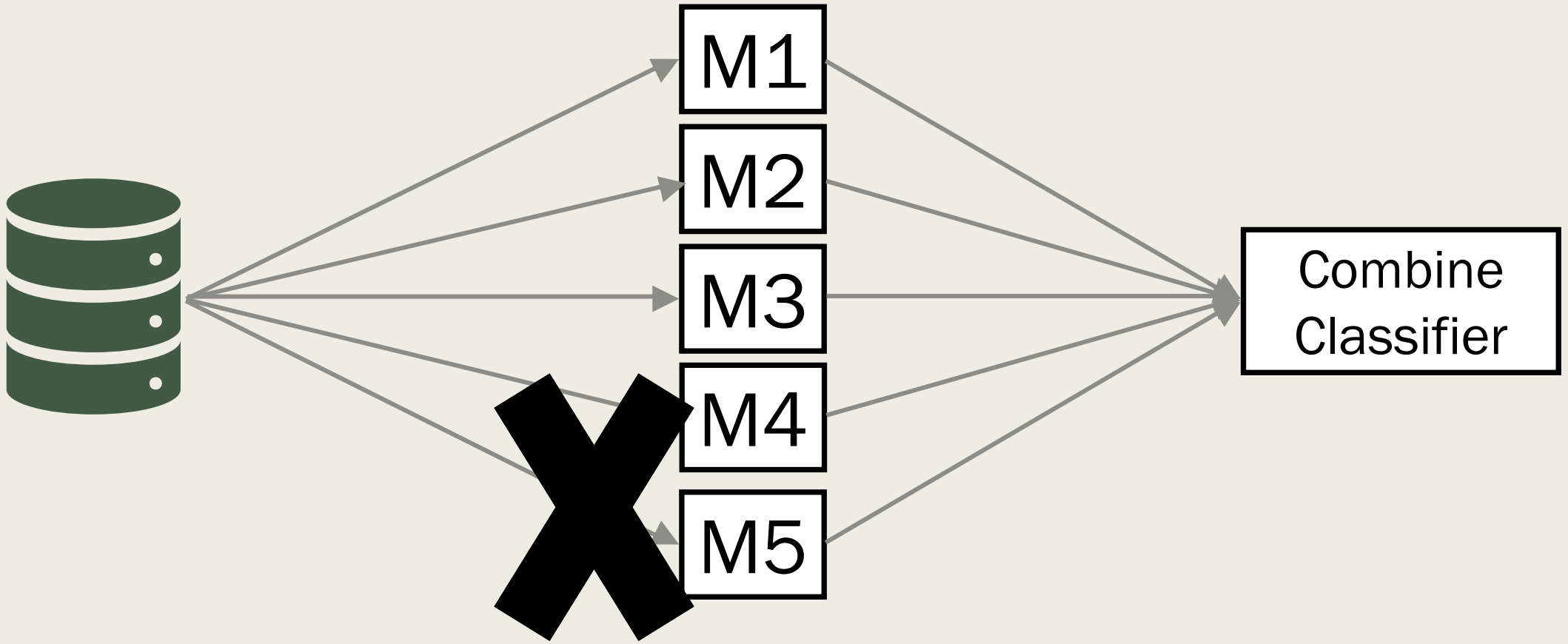
- XGBoost (Extreme Gradient Boosting)
- [Hands-on] XGBoost

BOOSTING

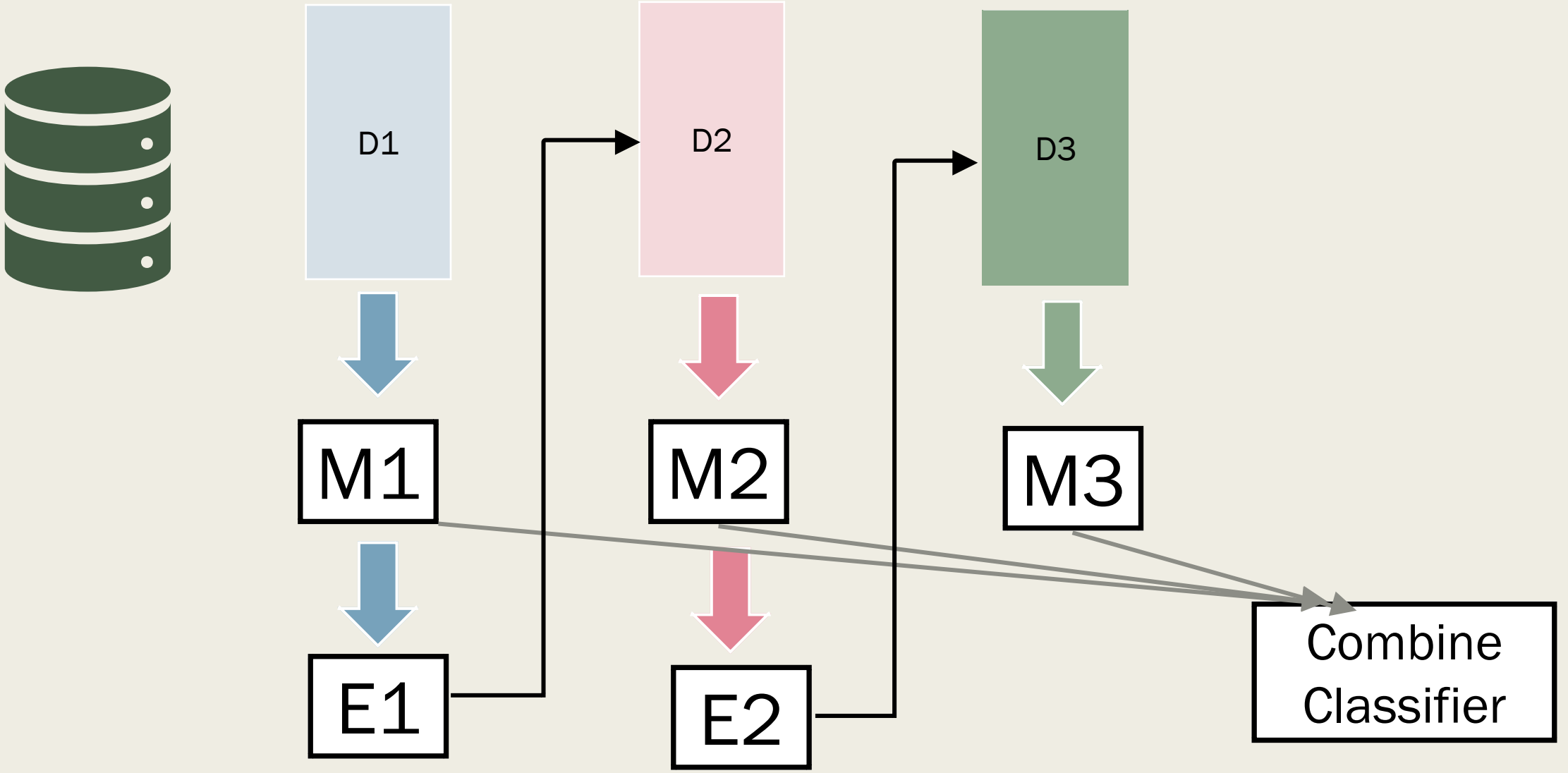
By Ankit Mistry

Boosting

- Do we use all training data on every model
- What is M1, M2, M3 …
- How to train all model
- How to combine output from all model



Boosting



By Ankit Mistry

Boosting

- Do we use all training data on every model
  - *Based on Sampling, Mistake*
- What is M1, M2, M3 …
  - *Same weak learner ML Algorithm – Decision Tree*
- How to train all model
  - *Train in Sequential*
- How to combine output from all model
  - *Any strategy based on Classification or Regression problem*
- Computational Time also very High

- Adaptive Boost
- Gradient Boost

ADABOOST

By Ankit Mistry

Adaboost

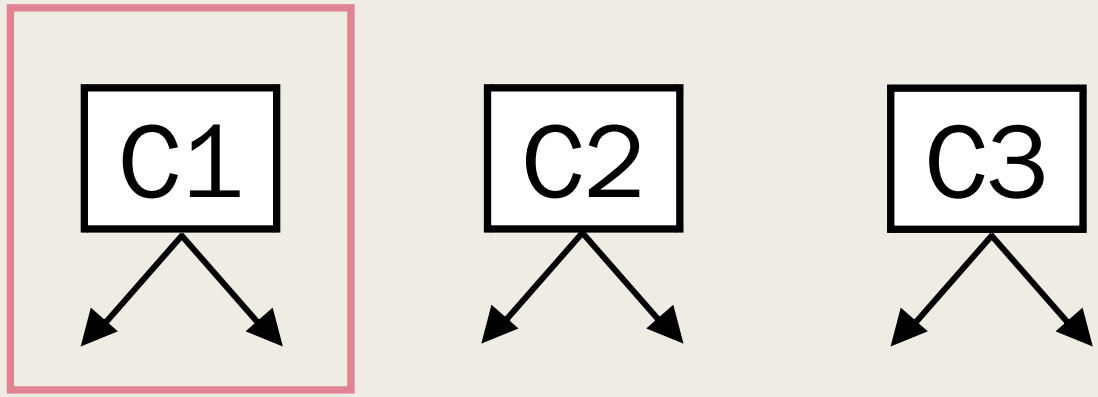
1. Stump Creation
2. Calculate Error and Stump Performance
3. Update Weight
4. Repeat Above 3 steps number of model you want to create as weak learner.
5. Combine classifier

Adaboost

C1	C2	C3	
			R1
			R2
			R3
			R4
			R5
			R6
			R7

Weight
1/7
1/7
1/7
1/7
1/7
1/7
1/7

C1	C2	C3	
			R2
			R3
			R6
			R7



- Calculate Entropy or Gini index
- Select Minimum Entropy Stump

C1	C2	C3	
			R3
			R6

Stump Performance

C1	C2	C3	
			R1
			R2
			R3
			R4
			R5
			R6
			R7

Weight
1/7
1/7
1/7
1/7
1/7
1/7
1/7

C1	C2	C3	
			R1
			R2
			R4
			R5
			R7

C1	C2	C3	
			R3
			R6

- Total Error = Sum of weight for Wrongly Classified Samples

$$= \frac{1}{7} + \frac{1}{7} = \frac{2}{7}$$

- Performance of Stump $= \frac{1}{2}\log\left(\frac{1 - Total\ Error}{Total\ Error}\right)$

$$= \frac{1}{2}\log\left(\frac{1 - 2/7}{2/7}\right)$$

$$= 0.458$$

By Ankit Mistry

Update Weight

Performance of Stump = 0.458

	Weight	Updated Weight	Norm Weight
R1	1/7	0.09	0.10
R2	1/7	0.09	0.10
R3	1/7	**0.22**	**0.25**
R4	1/7	0.09	0.10
R5	1/7	0.09	0.10
R6	1/7	**0.22**	**0.25**
R7	1/7	0.09	0.10

SUM = 0.89

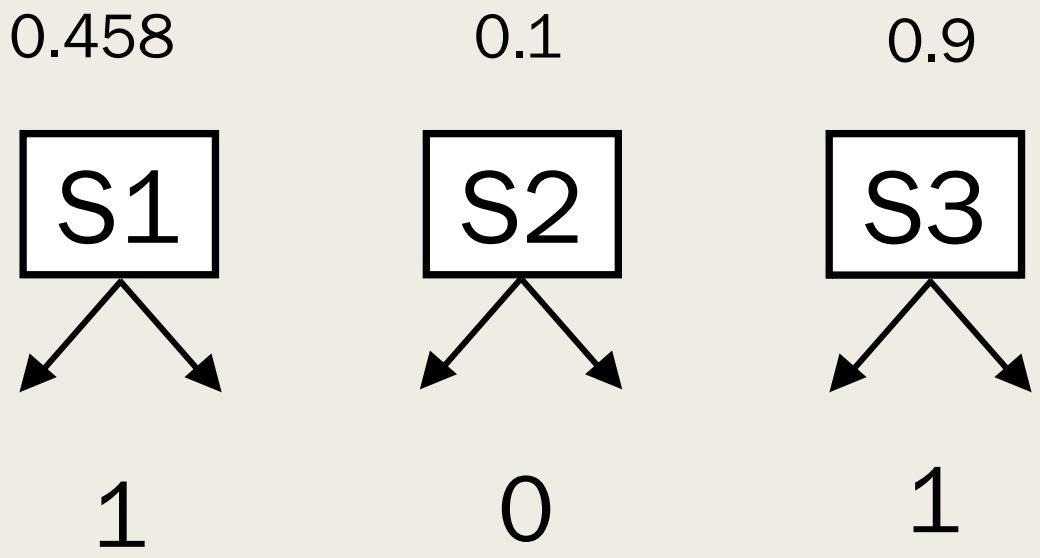
For Wrong Classification

$$New = Old * e^{stump\ performance}$$
$$= \frac{1}{7} * e^{0.458} = 0.22$$

For Correct Classification

$$New = Old * e^{-stump\ performance}$$
$$= \frac{1}{7} * e^{-0.458} = 0.09$$

Combine Classifier



Class 1 Support = 0.458 + 0.9 = 1.358

Class 0 Support = 0.1

1

(HANDS-ON) ADABOOST

By Ankit Mistry

GRADIENT BOOSTING

By Ankit Mistry

Gradient Boosting

1. Build Model and Make Prediction
2. Calculate Error (Residual) and Next target is error.
3. Build next model with target is error and Make Prediction
4. Update Prediction of earlier model by adding error.
5. Repeat above 3 steps number of Model you want to generate.

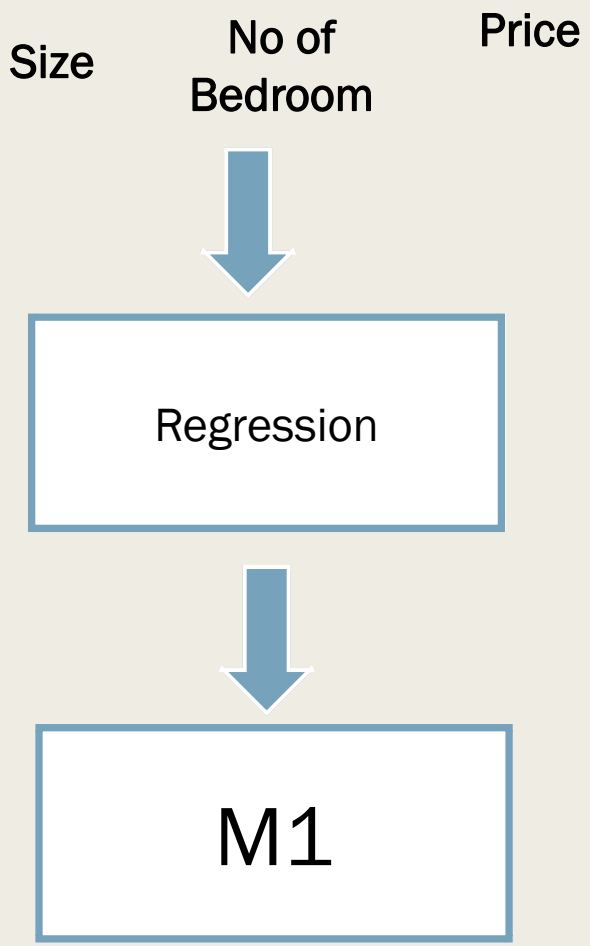
1. Build Model and Make Prediction

Gradient Boosting

Step 1

	Size	No of Bedroom	Price
1	2114	3	359
2	1700	3	229
3	1480	3	379
4	1294	2	212
5	1563	3	316
6	2420	2	365
7	1494	2	410

M1 - Price
312
250
370
260
300
352
390



Calculate Error (Residual) and Next target is error.

Gradient Boosting

Step 2

	Size	No of Bedroom	Price	M1 - Price	Error
1	2114	3	359	312	47
2	1700	3	229	250	-21
3	1480	3	379	370	9
4	1294	2	212	260	-48
5	1563	3	316	300	16
6	2420	2	365	352	13
7	1494	2	410	390	20

Error is now New Target

Build next model with target is error and Make Prediction

Gradient Boosting

Step 3

	Size	No of Bedroom	Price	M1 - Price	Error	Pred - Error
1	2114	3	359	312	47	20
2	1700	3	229	250	-21	-10
3	1480	3	379	370	9	3
4	1294	2	212	260	-48	-30
5	1563	3	316	300	16	5
6	2420	2	365	352	13	10
7	1494	2	410	390	20	11

M2

Error as Output

Update Prediction of earlier model by adding error.

Gradient Boosting

Step 4

	Size	No of Bedroom	Price	M1 - Price	Pred - Error	M2 - Price
1	2114	3	359	312	20	332
2	1700	3	229	250	-10	240
3	1480	3	379	370	3	373
4	1294	2	212	260	-30	230
5	1563	3	316	300	5	305
6	2420	2	365	352	10	362
7	1494	2	410	390	11	401

Go back to step 2 – again calculate error
(Updated price – Actual Price)

By Ankit Mistry

Gradient Boosting

1. Build M1
2. Build M2 = M1 + lambda * Model(error1)
3. Build M3 = M2 + lambda * Model(error2)
4. So on…

[HANDS-ON] GRADIENT BOOSTING

Gradient Boosting Tree

By Ankit Mistry

XGBOOST

Extreme Gradient Boosting Tree

By Ankit Mistry

XGBoost

- Extreme Gradient Boosting Tree
- It is not different algorithm but works on same principal used in GBM.
- Sequential model creation with keep on reducing Error.

Feature of XGBoost

- Faster Execution Speed
- Better Performance
  - *Many Machine Learning competition won by XGBoost*
- Parallel Processing.
- XGBoost can handle missing value
- It has Built–in cross validation , Regularization
- It Can do incremental training. (For Large Dataset which Can not fit in memory)

[HANDS-ON] XGBOOST

Extreme Gradient Boosting

By Ankit Mistry

SECTION 6

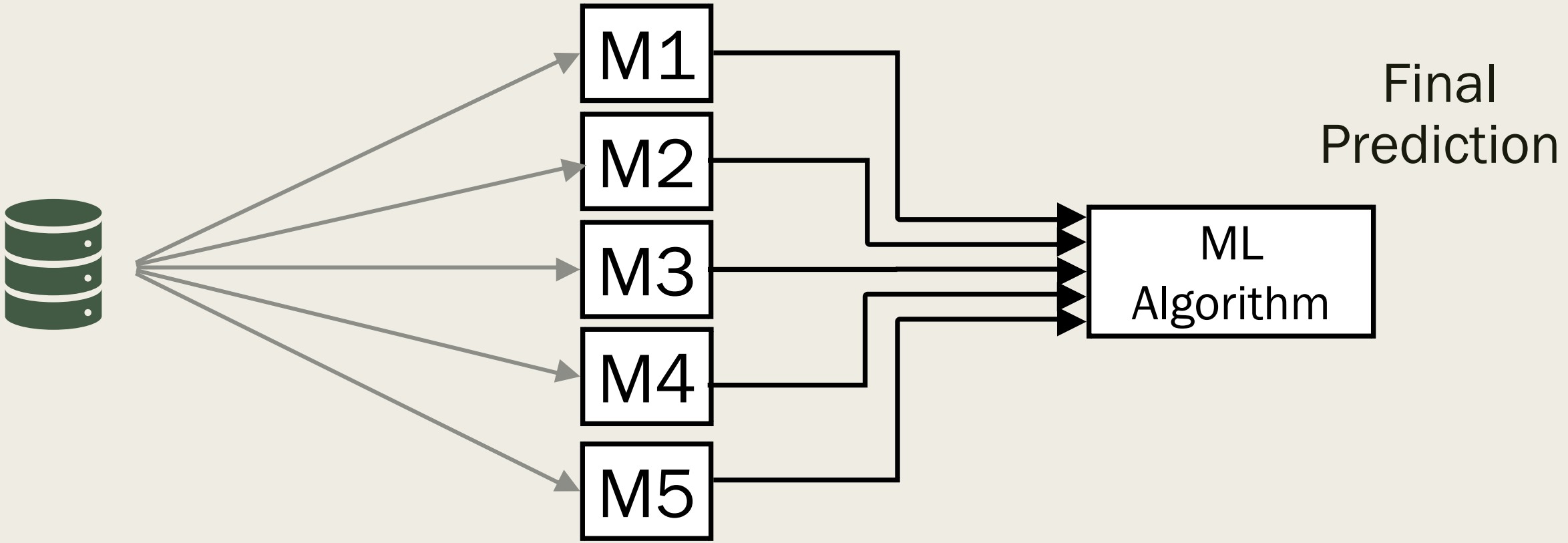
Section 6

- Model Stacking
- [Hands-on] Model Stacking

MODEL STACKING

By Ankit Mistry

Model Stacking



By Ankit Mistry

[HANDS - ON] MODEL STACKING

By Ankit Mistry

Model Stacking