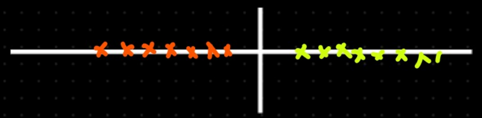
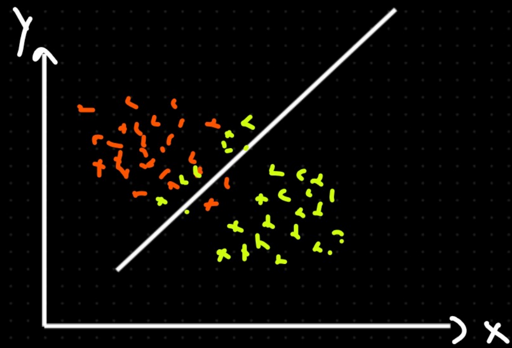
SVM kernels

function

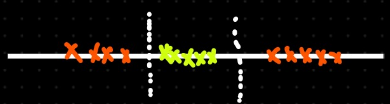
1. Regression
2. Classification

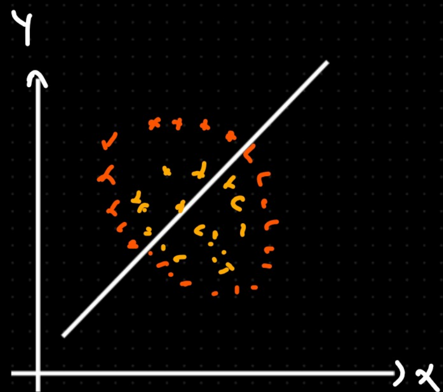
Linear Seprable data





Non-linear Seprable data





Transformation

Data → Transformed data

1. Polynomial
2. Sigmoid
3. RBF

} Convert data into higher Dimension

{lower Dim → higher Dim}

{
1D → 2D
2D → 3D
3D → 4D
}

transform this $\boxed{x}$ variable ⇒ x^2

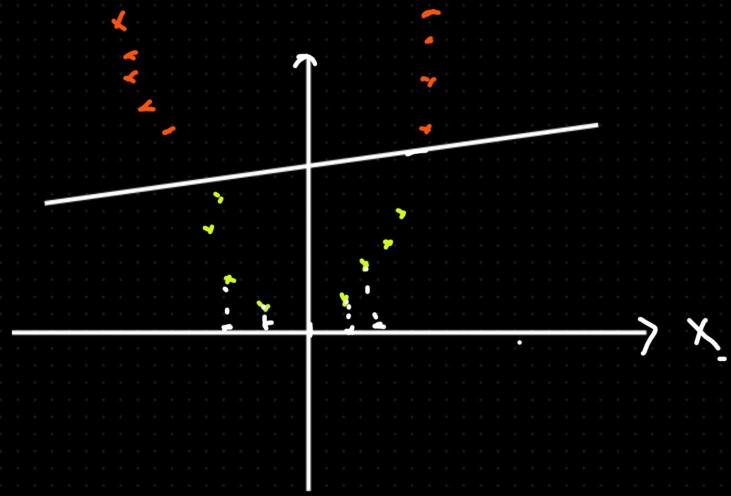
x_1 x_2

$\boxed{X \; x \; x}$

1D

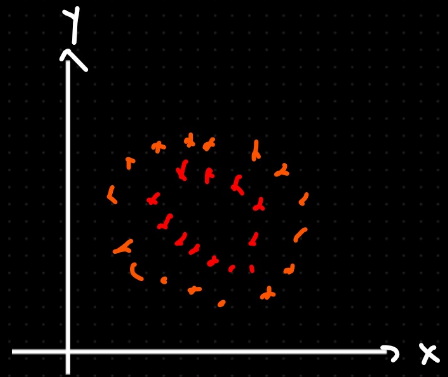
$X = \boxed{y = x^2}$

$Y = x^2$



$\rightarrow \boxed{Y = f(x) = x^2}$

Kernel



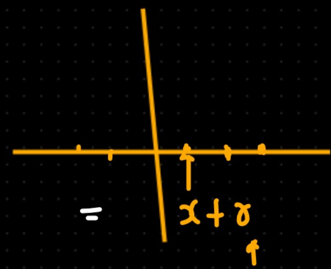
Polynomial kernel

$$(x_1 x_2 + r)^d \qquad d = 1,\ r = 0$$

$$\Downarrow$$

$$(x_1 x_2 + 0)^1$$

$$\Rightarrow (x^2)$$



$f(x_1, x_2) \Rightarrow$ Polynomial kernel $\Rightarrow (x_1 x_2 + r)^d$

$\Downarrow$

function

$X_1 \ X_2 \ Y \Rightarrow$ Binary class

OVR

RBF kernel $= f(x_1, x_2) = \exp^{-\left(\frac{\|x_1 - x_2\|}{2\sigma^2}\right)}$

fn

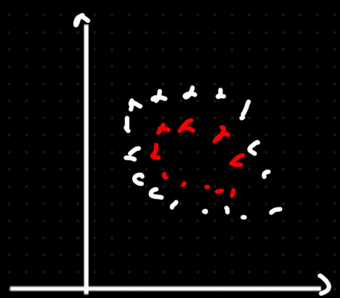
$\Rightarrow (x_1 x_2 + r)^d \Rightarrow \left[x_1^2 , \ x_1 \cdot x_2 , \ x_2^2 \right]$

$\Downarrow$

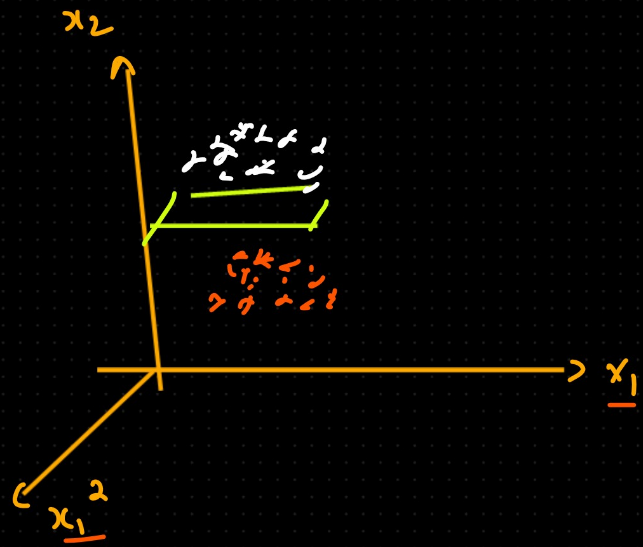
2D

$X_1 X_2$

$X_1 \quad X_2 \quad Y \quad x_1^2 \quad x_2^2 \quad x_1 x_2$



$\Rightarrow$



- linear reg → Ridge, Lasso, Elastic } Regularization
- log reg
- SVM
- DT

Reducing overfitting

low bias + high variance

generalization - Reg

→ L1, L2, L1+L2

↓

Cost Fn

Ensemble technique => More than one model

or

compound technique

- Bagging

SVC DT

Multiple model Parallaly

custom model

Decision tree → Random forest

- Boosting

multiple model sequentially

Custom

- Ada Boosting
- Gradient Boosting -
- XGB -

= Decision tree

Catboost

- Stacking

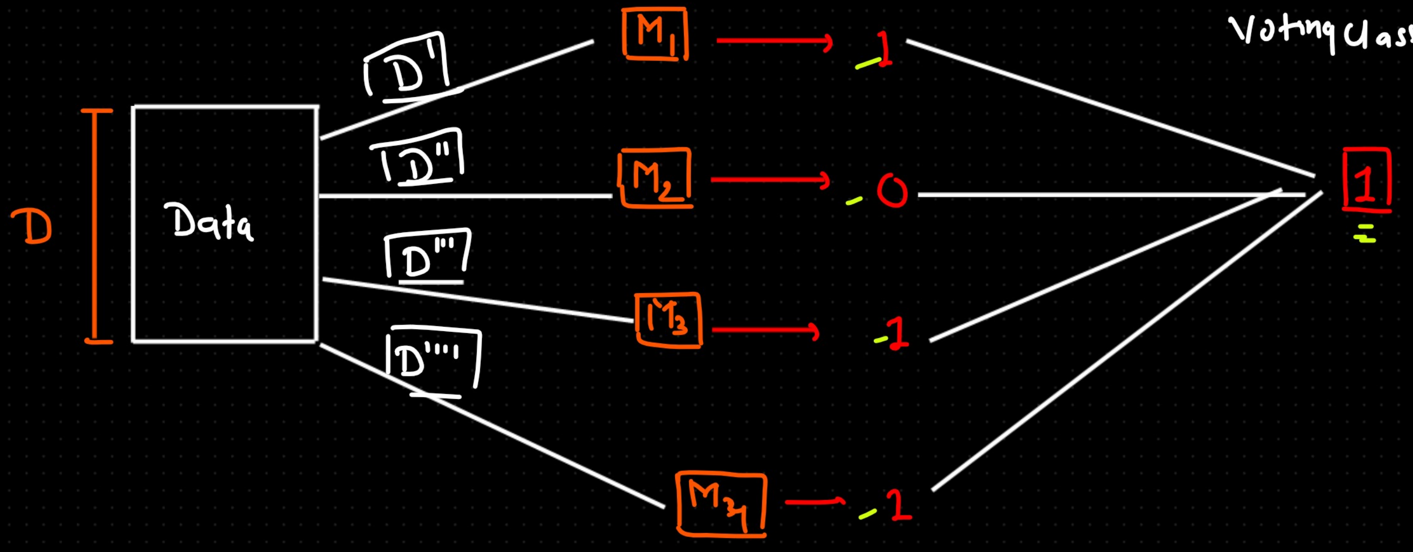
multiple model in 2 stages

SVC, DT, linreg

Bagging

Decision tree | Naive bayes | Logreg | SVM

Voting classifier (majority vote)



0 , 1
↓ ↓
→ 1 3

→ 0,1,1,01

2,2

Bootstrapping

Aggregation

Bagging ⇒ Bootstrap + Aggregation

↓ Sample

↓ majority vote

Regression → Avg of all the output

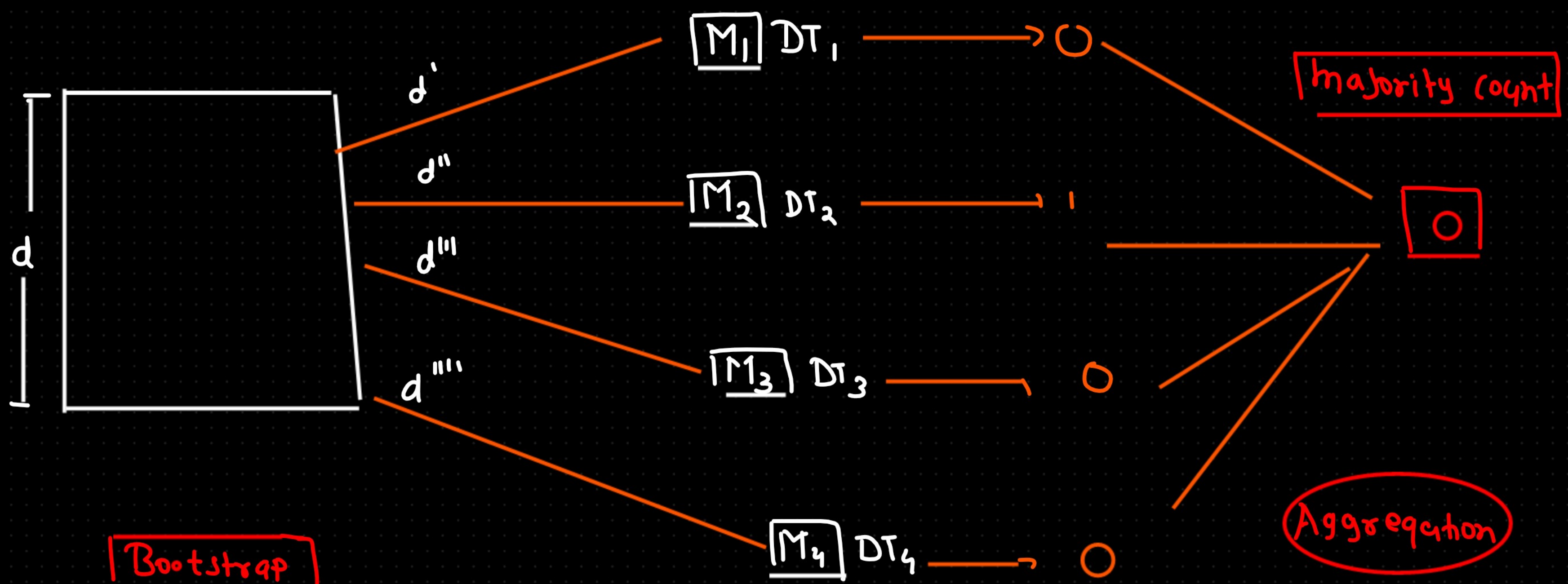
Random forest ⇒ the base learner will be Decision tree.

$d' < d$

$d'' < d$

$d''' < d$

$d'''' < d$



Bootstrap

D

	x_1	x_2	x_3	x_4	x_5
→	1	5	5	12	12
	2	6	10	14	18
	3	7	11	15	15
→	4	5	12	16	20

$d' = [1, 4, x_1, x_4]$

$d'' = [1, 2, x_2, x_5]$

⋮

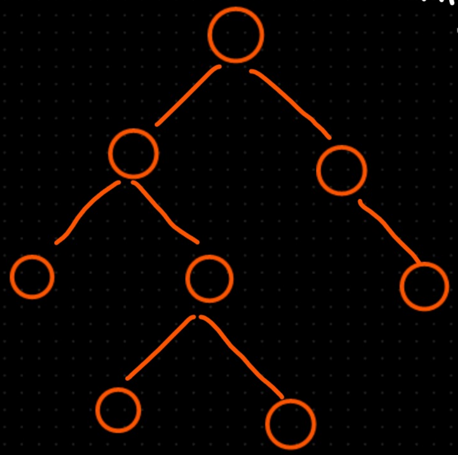
$d' / d'' / d''' / d''''$ ⇒ Subset of the data ⇒ Random Sample

(Row sample) + (Column sample)

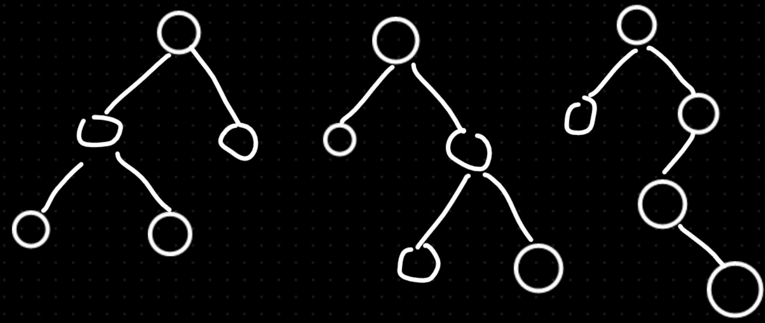
= Bootstrap Data

Why we should Random forest instead of Decision tree?

↳ multiple tree



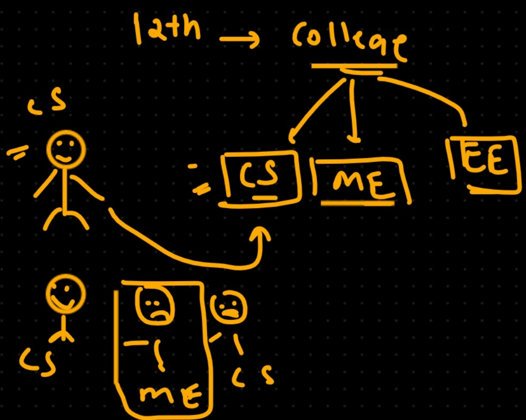
⇒

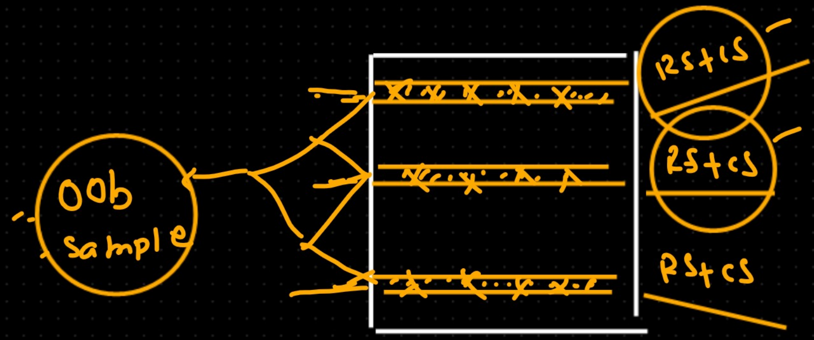


(Generalized model)

till complete depth → Overfitting ⇒ training Acc ↑↑↑↑ , testing Acc ↓↓↓↓ } $\boxed{\text{low bias, high variance}}$ ⇒ $\boxed{\text{low bias, low variance}}$

1. Reduce the overfitting
2. Better performance
3. Robustnen
4. feature importance





$\boxed{OOB = True}$