

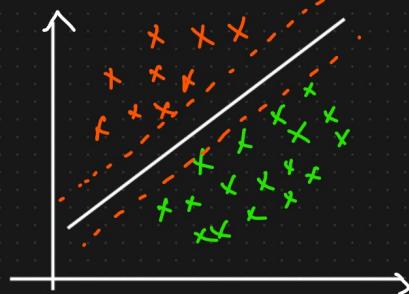
## Agenda

- ① SUM Kernels → Practicals
- ② Bagging & Boosting Algorithms → Random Forest, Adaboost
- ③ Pick up any problem statement
- ④ ROC and AUC Curve } → Performance Metrics

Applied Data  
// Scientist

Amazon, Microsoft

### ① SVC & SVR



Cost function

Minimize  
( $w, b$ )

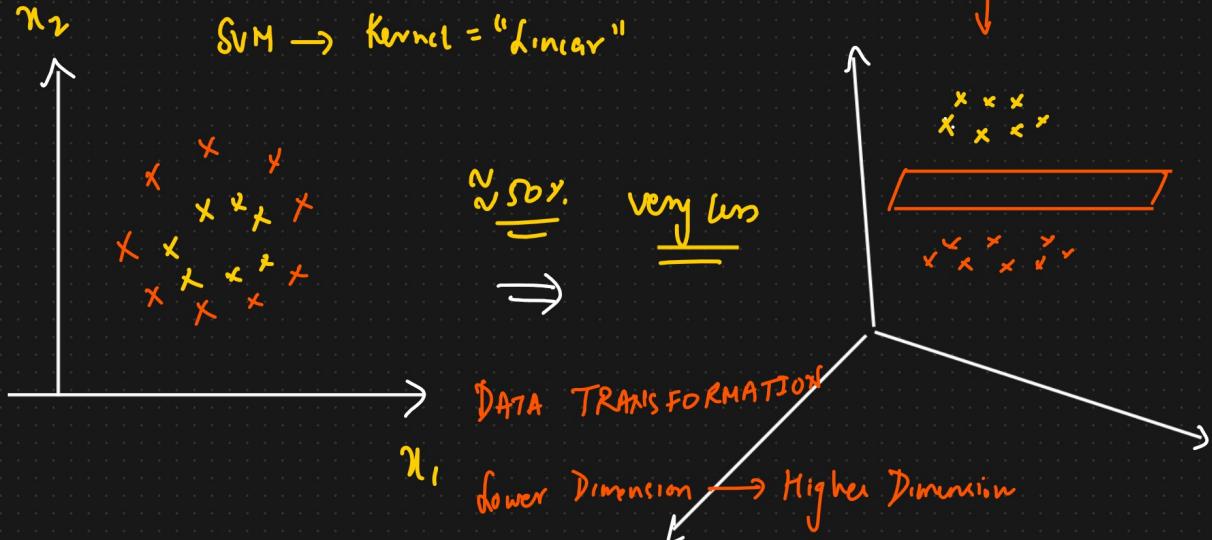
$$\frac{\|w\|}{2} + C \sum_{i=1}^n \xi_i$$

Constraint

$$y_i w^T x_i + b \geq 1$$

→ for all correct points.

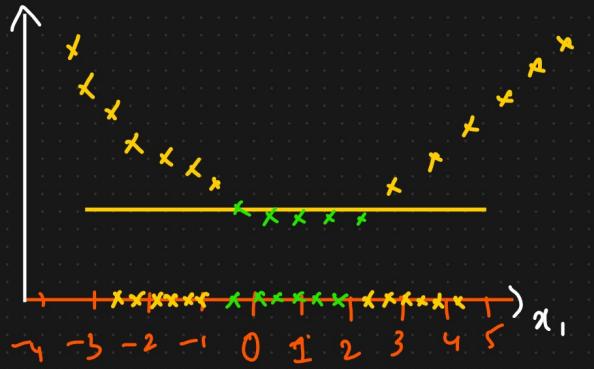
SVM → Kernel = "Linear"



PCA = High Dimension → Low Dimension

① Polynomial ✓

2D → 3D



② RBF ✓      ③ Sigmoid Kernel ~~sign~~  
 1 dimension → 2 dimension

Applying one mathematical  
formula  
=====

$$X = \{x_1, x_2\}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$y$$

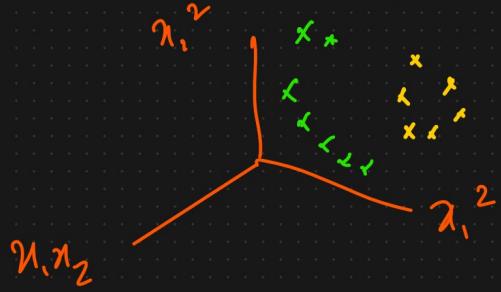
$$w^T x$$

$$d=2$$

$$\text{Polynomial Kernel} = \left( \underline{\underline{x^T \cdot x}} + 1 \right)^d$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [x_1, x_2] = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}$$

$$x_1 \quad x_2 \quad \begin{bmatrix} x_1^2 & x_1 x_2 & x_2^2 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{linear} \\ \text{Hyperplane} \end{bmatrix}$$



Sklearn Kernel = "poly"

② RBF Kernel

{Panasonic}



$$x_1 = \{1, 2, 3, 1\}$$

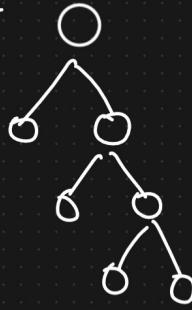
$$x_2 = \{5, 6, 7, 8\}$$

$y_{stack}(x_2)$



④ Decision Tree  $\Rightarrow$  Overfitting

DT

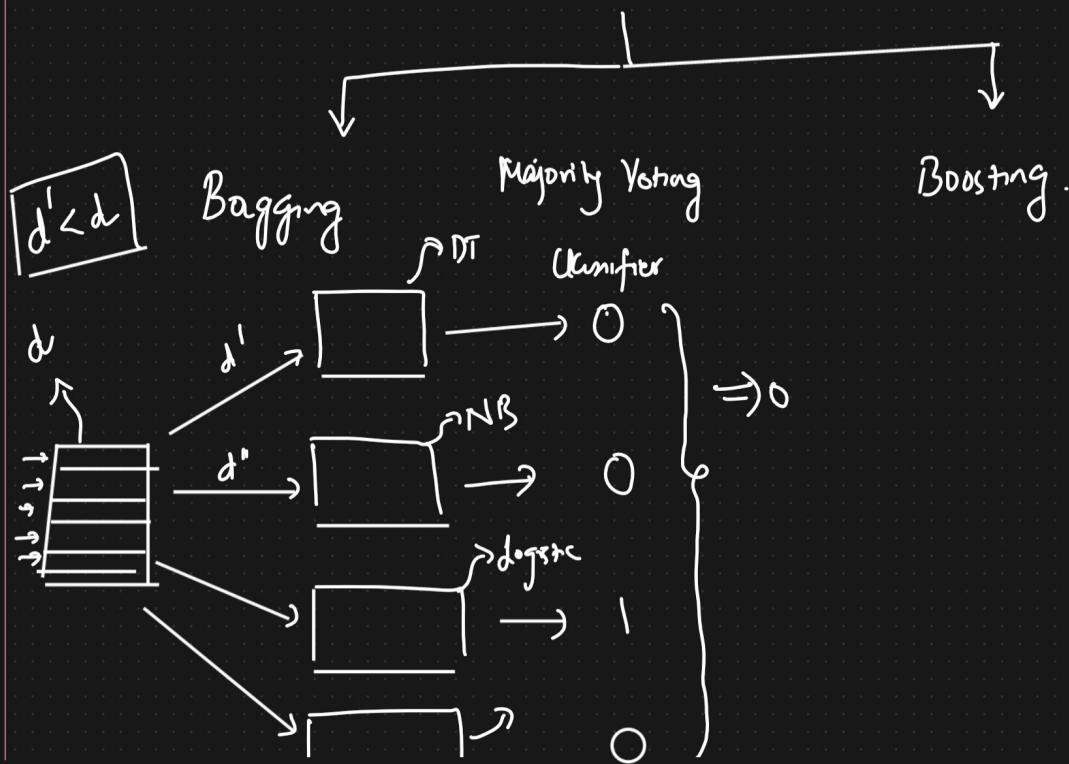


Overfitting

$\Rightarrow$  {  
low Bias  
High Variance}

low Variance {Ensemble Techniques}

## Ensemble Technique

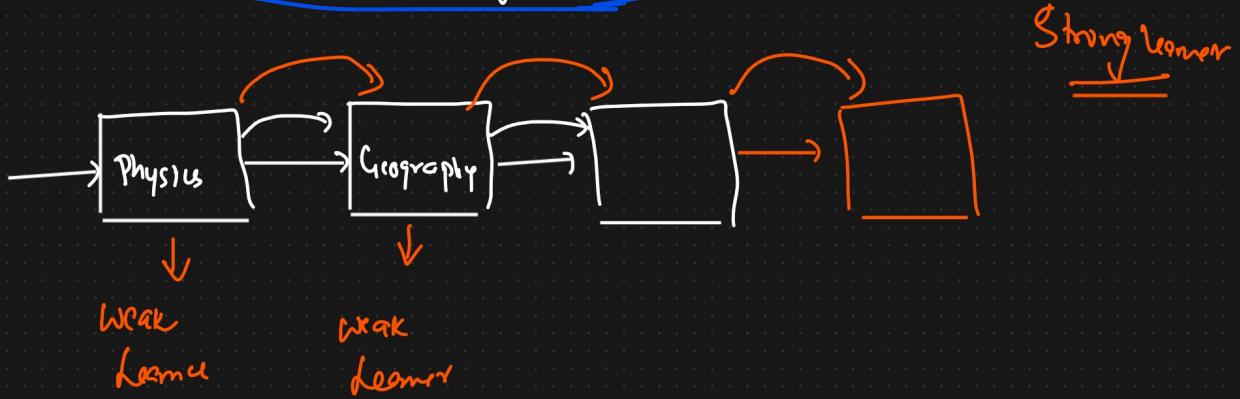


Avg(O/P)

Boosting → KBC

{ AdaBoost  
Gradient  
XGBoost }

→ Question → History Subject → Sequential



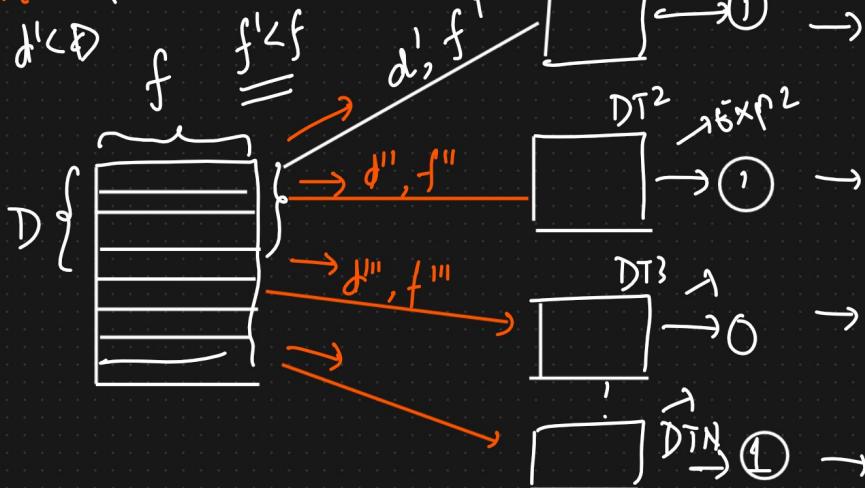
Bagging

① Random Forest Classification & Regression

$f^l$  = subset of features

$d^l$  = subset of datapoints  
Classification

Random Forest  
Test data



Boosting

① AdaBoost

② CatBoost

③ Gradient Boost

④ Xg Boost

} {

① or ②

Mean( ) ⇒ Answer  
=

{  
 do Bias  
 &  
 low variance

---

## Performance Metrics

ROC and AUC Curve

[0, 0.2, 0.4, 0.6, 0.8, 1]

Healthcare Problem

Threshold = 0.5

$y$	$\hat{y}_{\text{pred}}$	$\hat{y}(0)$	$\hat{y}(0.2)$	$\hat{y}(0.5)$
1	0.8	1	1	1
0	0.96	1	1	1
1	0.4	1	1	0
1	0.3	1	1	0
0	0.2	1	0	0
1	0.7	1	1	1

$$\text{TPR} = \frac{2}{2+2} = 0.5$$

$$\text{FPR} = 0.5$$

$$\text{TPR} = \frac{4}{4+0} = 1$$

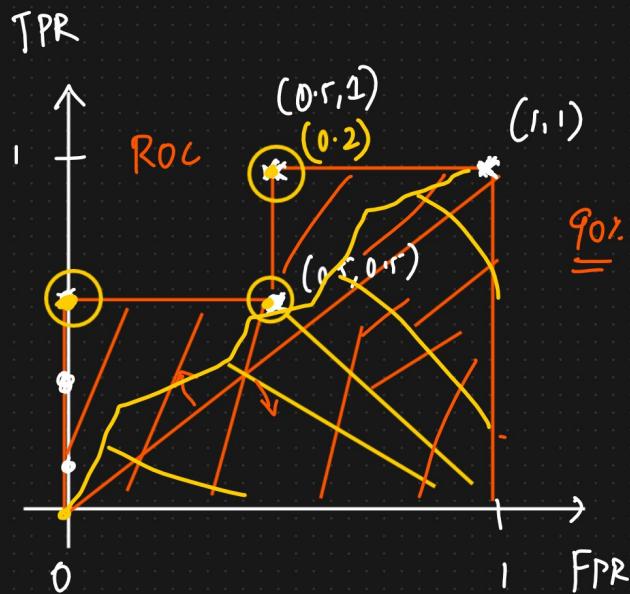
$$\text{FPR} = \frac{1}{(1+1)} = 0.5$$

Probability

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{4}{4+0} = \boxed{1}$$

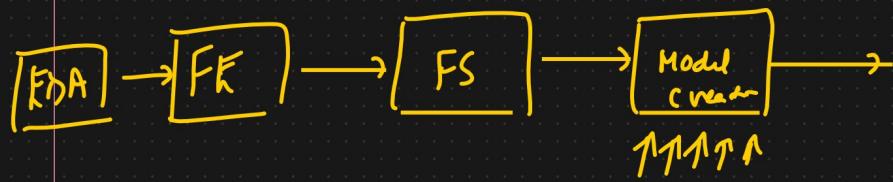
$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{2}{2+0} = \boxed{1}$$

Domain Expertise  $\Rightarrow$  TP

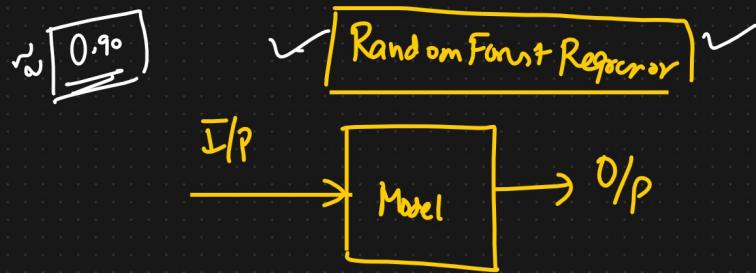


AUC

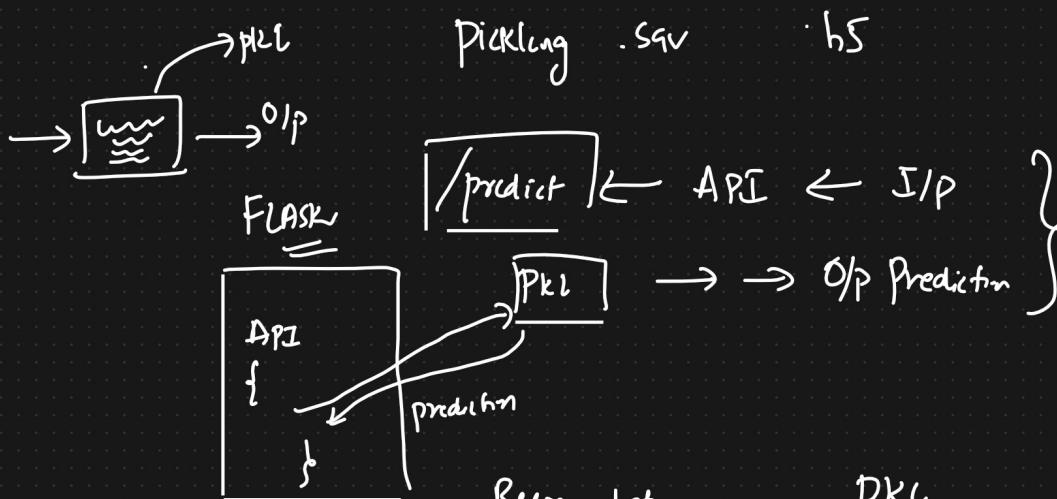
Threshold and Model performs well.



How should be  
use this model



## Serialized Object format



{ → Model pred  
} ← q<sub>pr</sub>  
Postman



{ load the model  
{ }