

→ Decision Tree:

→ Log. Reg.

→ Lin Reg.

→ KNN:

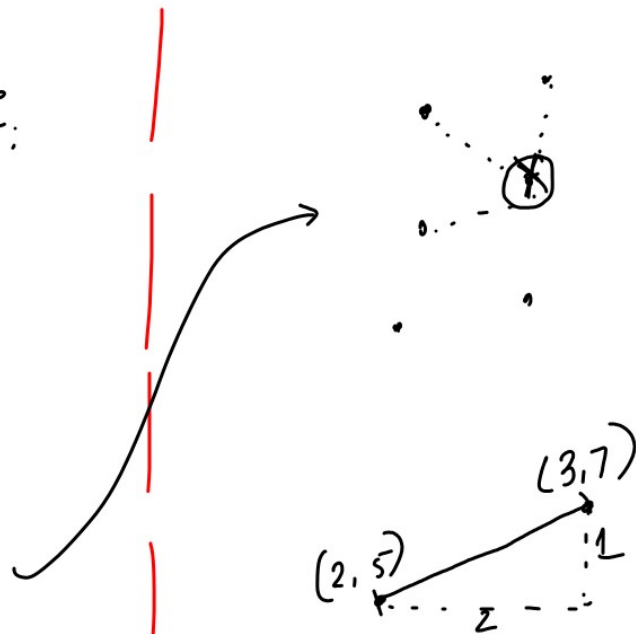
→ Complete KNN

→ Ensemble Methods:

→ Bagging.

→ Random Forest.

→ Boosting. → Ada boost
→ Gradient boost



$$d_c = \sqrt{2^2 + 1^2} \\ = \sqrt{5}$$

✓ → Euclidean Distance

$$d = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}$$

✓ → Manhattan distance

$$d = \sum |x_i - y_i|$$

→ Minkowski Dist.

$$d(x, y) = \left(\sum |x_i - y_i|^p \right)^{1/p}$$

$$p = 1 \quad (x_2 - x_1) + (y_2 - y_1)$$

$$p = 2 \quad \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$p = 3 \quad \sqrt[3]{(x_2 - x_1)^3 + (y_2 - y_1)^3}$$

Ensemble Models

	M_1	M_2	M_3	M_u	...	
1	✓	✓	✗			✓
2	✓	✗	✗			✗
3	✓	✗	✓			✓
4	✓	✓	✓			✓
5	✗	✓	✗			✗
6	✗	✗	✓			✗
7	✗	✓	✗			✗
8	✗	✗	✓			✗
9	✓	✓	✓			✓
10	✓	✓	✓			✓
	60%	60%	60%			75%

Accuracy =
 Precision =
 Recall =

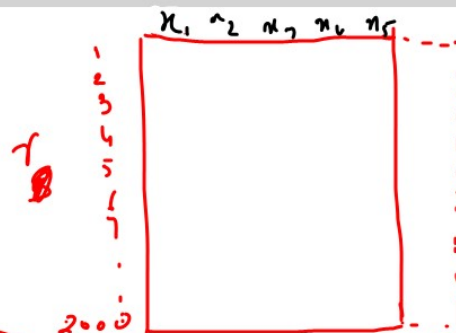


New



~~bagging~~

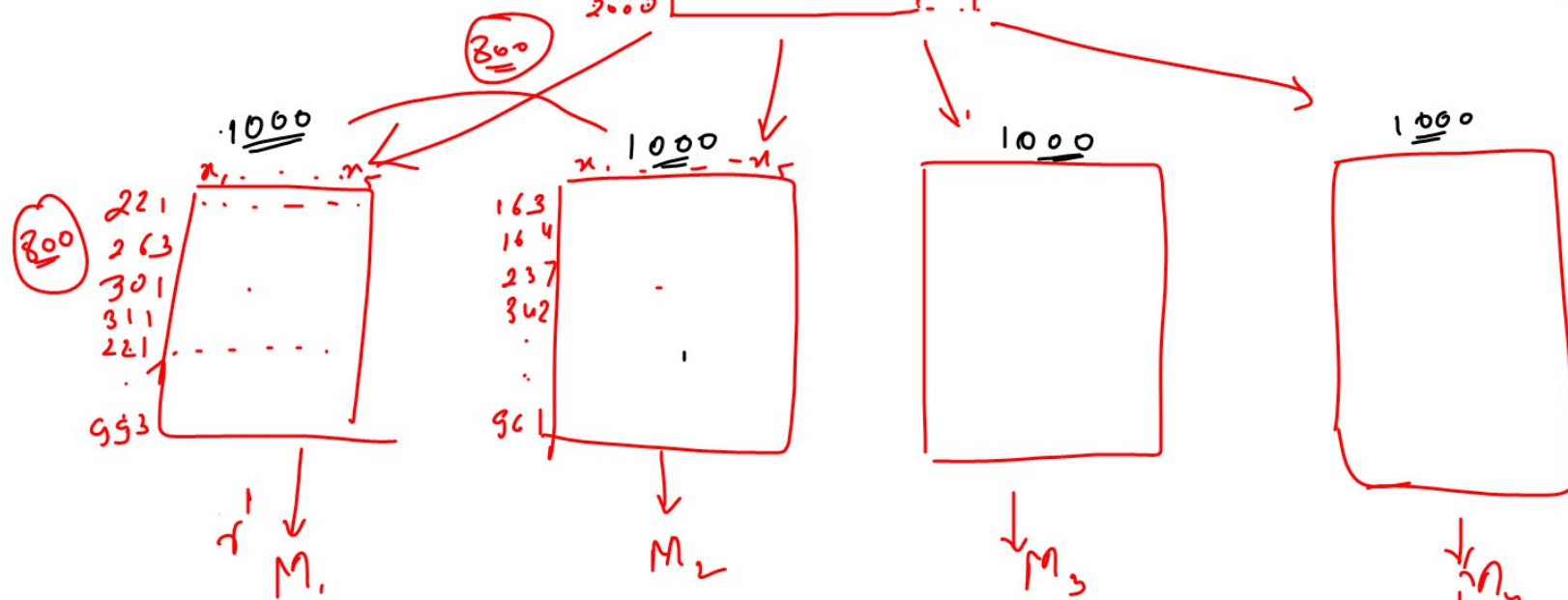
Subsets:



$$\gamma^1 = \gamma$$

$$\gamma^1 < \gamma$$

$$\frac{2000}{1066}$$



→ Correlated
(Independent)
(min Similarity)
→ All of them
must not be
weak.

45%
=:
50%
=:
55%

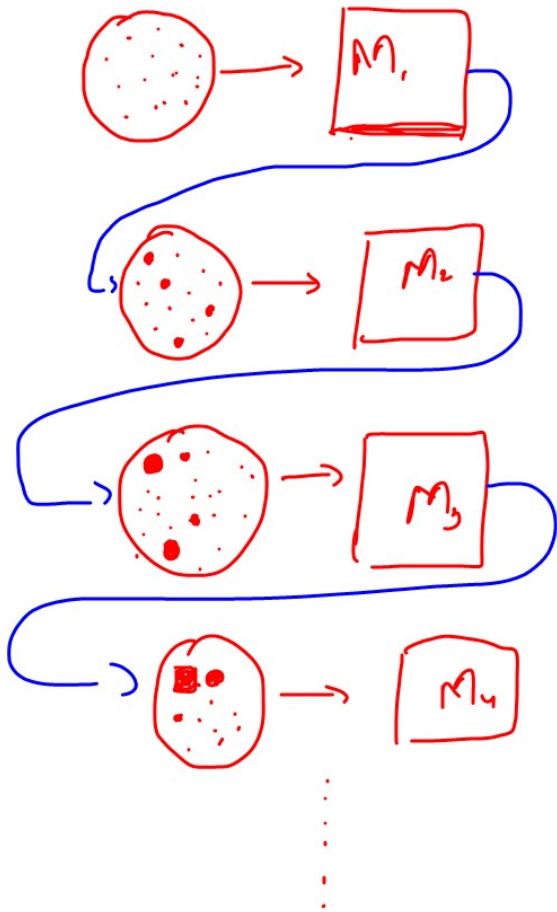
Boosting

Adaboost

Gradient boost.

XG boost.

Random forest.



New



Random forest:

x_1, x_2, \dots, x_{10}

$A, B, C, D, E, F, G, H, I, J, Y$

$m \leq M$

$m=5$

$2000 \times 10 \times 5$

2000

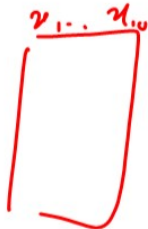
$R, G, H, I, J \Rightarrow Y$

A, B, C, E, G

B, C, D, E, F

1150

|||||

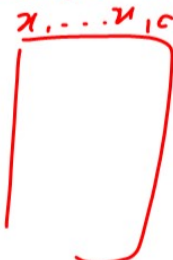


$M_1 = DT$

1150



$M_2 = DT$



$M_5 = DT$



$M_3 = DT$



New

