21/3/24

JUNE 2023

SUN	MON	TUE	WED	THU	FRI	SAT
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

MAY'23
FRIDAY
26
21st Week • 146-219

46.

SVM

SVM ⇒ SVR ⇒ Regression

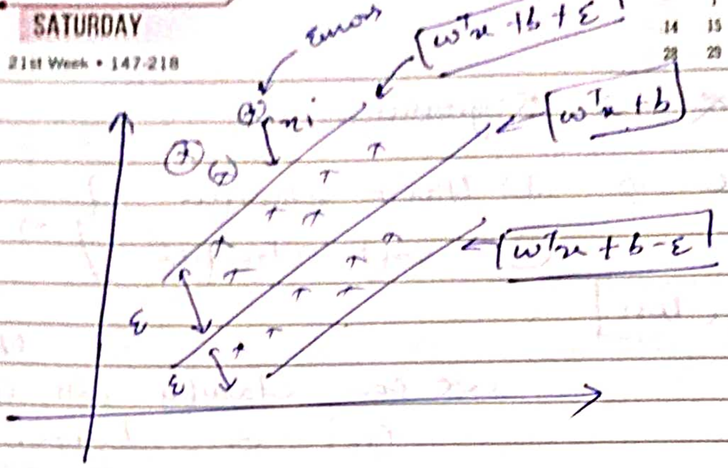
SVC ⇒ ① Hard classifier
② soft classifier
→ linear separable data
↓
the data which we can classify with the single line or hyperplane
↳ 3D or 4D.

loss fun ⇒ Hinge loss

non - linear separable data.

Kernal concept
↓
formula
↓
transform higher dim
↓
then we classify the data

Kernel → ① RBF Kernel
↓ → ② polynomial kernel
③ sigmoid kernel



linear for regression is support vector

Cost fun $\Rightarrow \min_{w,b} \frac{\|w\|}{2} + C \sum_{i=1}^{n} \varepsilon_i$

$x_i = \varepsilon_i$ eeta

ε = Epsilon

Constraint

$|y_i - w^T x_i| \le \varepsilon + \varepsilon_i$ = |Data point − plane|

ε = marginal Error = value $\le \varepsilon_i + \varepsilon$

ε_i = Error above the margin

linear Kernel $|w^T x + b_0|$

polynomial Kernel $|(w^T x + c)^{degree}|$

* sigmoid Kernel $|\tanh(\alpha \times w^T x + c)|$

RBF Kernel $|\exp(-gamma \times |x - y|^2)|$

Decision Tree

In ml we are having different approaches for solving a data.

Decision tree ⇒ ① linear algebra → linear reg ✓
→ SVM ✓

② Probabilistic → logistic reg ✓
→ Naive Bayes.

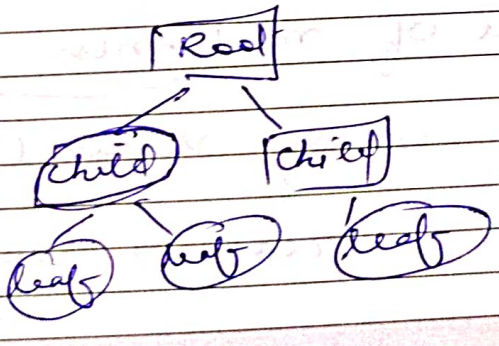
③ clustering → Kmeans
→ DB scan
hierarchical

④ tree → decision tree
→ random forest.
AB
GB
XB

⑤ Neural network → RNN
→ CNN
→ ANN

In tree based approach we have decision tree.

decision tree ⇒ it is based on condⁿ.



29 MAY'23 MONDAY

22nd Week • 149-216

To build decision tree we generally use 3 algo.

1. ID3 algo (iterative dichotomiser)
2. CART (classificatn & reg tree)
3. C 4.5 (extension of ID_3)
4. CHID categorical → chi-square test

ID_3	CART
1. iterative Dicotomiser	1. classificatn & reg tree
2. entropy → informatn gain	2. Gini impurity, Gini coefficient, Gini index
3. multiclass	3. binary class
4. only works for the classificatn	4. works for reg and classificatn.

Entropy ⇒ measure of randomness → error.

used in ID3

Range of entropy value [0, 1]

If near to zero → less random or more robust data

If near to 1 → more random

The most profitless thing to make is an excuse.

$$\text{Entropy} = -\sum_{i=1}^{N} P_i \log_2 (P_i)$$

$i = 2 \Rightarrow$ yes / no

$$-P_y \log_2 (P_y) - P_N \log_2 (P_N)$$

entropy of 2 classes yes & no.

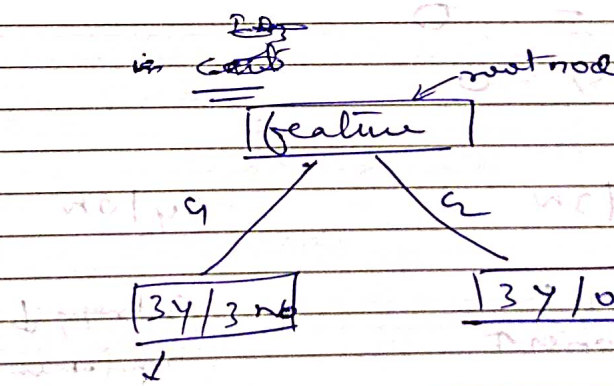
Entropy for 3 classes

$$-P_1 \log_2 (P_1) - P_2 \log_2 (P_2) - P_3 \log_2 (P_3)$$

* ⇒ In Entropy we find out best feature for the root.

Eg) Data

Feature	O/P
c_1	Yes
c_2	Yes
c_1	Yes
c_2	Yes
c_1	Yes
c_1	No
c_2	Yes
c_1	No
c_1	No



Entropy =

$$-\frac{3}{6} \log_2 (3/6) - 3/6 \log_2 (3/6)$$

[P.T.O]

$$= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - 1/2 \log_2(1/2)$$

$$= -1/2\left[\log(1) - \log_2(2)\right] - 1/2\left[\log_2(1) - \log_2(2)\right]$$

$$= -1/2\ [0 - 1]\ -1/2[0 - 1)$$

$$= -1/2\ (-1) - 1/2(-1)$$

$$= 1/2 + 1/2$$

$$= 1$$

Entropy = $H(S) = 1$

For 3yes | 0 no

$$-\frac{3}{3}\log_2(3/3) - \frac{0}{3}\log_2(0/3)$$

$$= -\log_2(1)$$

$$= 0$$

$H(S) = 0$

note

3y | 3N

Entropy ↑
randomness ↑

6y | 0N

Entropy ↓
randomness ↓

In this case we will choose 6yes | 0 no.

Entropy = 0

In 3ya/3NO we have 50 – 50% chance

interview

a) calculate entropy.

(2yes / 3no)

⇒ Total = 5.

$$-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right)$$

$$\Rightarrow -\frac{2}{5}\left[\log_2(2) - \log_2(5)\right] - \frac{3}{5}\left[\log_2(3) - \log_2(5)\right]$$

$$\Rightarrow -\frac{2}{5}\left[1 - 0.698\right] - \frac{3}{5}\left[0.477 - 0.698\right]$$

$$\Rightarrow -0.3396 \quad -0.6442$$

$$= -0.983 \qquad = 0.971$$

Gini Impurity used in cart.

$$1 - \sum_{i=1}^{N} (p_i)^2$$

i = class

two class ⇒ Y/N

$$= 1 - \left[p_Y^2 + p_N^2\right]$$

↓ less randomness or less feature impurity select that data.

3 yes | 3 no.

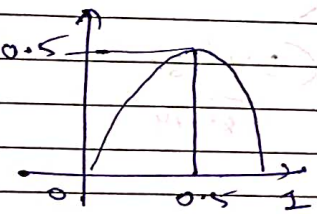
$$\left[1 - \left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right]$$

$$1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right]$$

$$1 - \left[\frac{1}{4} + \frac{1}{4}\right]$$

$$= 0.5$$

(50 - 50 situation)



3 yes | 0 no

$$1 - \left[\left(\frac{3}{3}\right)^2 + \left(\frac{0}{3}\right)^2\right]$$

$$= 0$$

Interview

Entropy

→ value of entropy will be from [0, 1]

→ it will be slow.

→ use for less dataset.

Gini Impurity

→ value of G.I will be from [0, 0.5)

→ it will be fast.

→ use for large dataset.

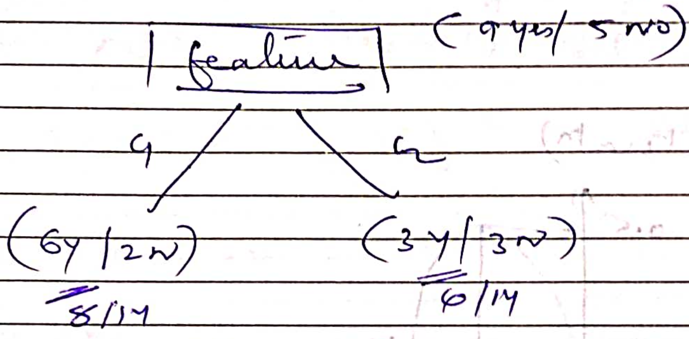
Information Gain

$$\text{Gain} = H(S) - \sum_{v\,(\text{value})} \frac{|S_v|}{|S|} H(S_v)$$

$H(S)$ = root entropy

$H(S_v)$ – child entropy

Q)



$H(S) \Rightarrow$ root feature

$= -P_y \log_2(P_y) - P_N \log_2(P_N)$

$= -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$

$= -(0.64)\log_2(0.64) - 0.35\log_2(0.35)$

$H(S) = 0.9402$

6Y/3N $\Rightarrow$ child entropy $H(S)$ L

$= -\frac{6}{8}\log_2(6/8) - 2/8\log_2(2/8) = 0.81$

3Y/2N $\rightarrow$ $H(S)$ R $= -\frac{3}{6}\log_2(3/6) - 3/6\log_2(3/6) = 1$

$$IG = 0.94 - \left[\frac{8}{14} \times 0.81 + \frac{6}{14} \times 1\right]$$

$$= 0.94 - [0.462 + 0.42]$$

$$= 0.0485$$

Conclusion :-

If I will get more info gain that feature or that split I will consider.