02/01/25 L2

12 x = 27 y } 6

3 { 15 x = 33 y

9 { 24 = 51 } - 18

12 { 36 = 75 } - 24

{ 2000

1964 → 3928

4003

18

36

42

~~2122~~

27

12 ~~2~~ 34 } 6

3 { 15 2 40 } 21

9 { 24 3 61 } 24

12 { 36 3 85

predictions

features

no of dimensions.

- Linear data / Eq - dependent
- Non-linear data / Eq - independent
- Slope remain const

Learning:

Assume a model $y = \theta_0 + \theta_1 x_1$ → Extract optimal values for parameters (θ_0, θ_1 from data)

↓

Learned model $y = 3 + 2x_1$

θ → parameters of a model

* Learning is about extracting parameters, using ML algo, from an assumed linear / non-linear model from which the data would have been generated.

ML Types:

Learning Problems.

1. Supervised Learning
   learn an input - output
   - classification - categorical output.
   - Regression - continuous output
2. Unsupervised Learning
   - Discover patterns in data
   - clustering - cohesive grouping
   - Association - frequent co-occurence

- not optimal but approximate

$$\frac{\Delta J}{\partial \theta_j} = (h_\theta(x) - y)\, x_j$$

$$\theta_j = \theta_j + \alpha \sum_{i=1}^{n} \left(y^{(i)} - h_\theta x^{(i)} \right) x_j^{(i)}$$

Least Mean Square update (LMS) / Widrow-Hoff Learning rule.

Repeat until convergence. (do for each j)

* Stochastic Gradient (Probabilistic) Descent (SGD).
- Batch Gradient Descent 2 Looking in Entire Training data

Convex function $\rightarrow J(\theta)$

x_1	x_2	y
1	2	3
2	1	4
3	3	5

25/10/21

$\theta_0 = \theta_1$

$x_0 = 1$

$x_0 = 1$

$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2$

$\theta_0 = \theta_1 = \theta_2 = 0$

$h_\theta = 0$

$\cancel{J(\theta) = \frac{1}{2}}$

S1

$SE = (0 - 3)^2 = 9$

S2

$SE = (0 - 4)^2 = 16$

S3

$SE = (0 - 5)^2 = 25$

$SSE = (9 + 16 + 25)$
$= 50$

$J(\theta) = \frac{1}{2} \times 50$
$= 25$

$\theta_0 = \theta_1 = \theta_2 = 0.1$

S1

$SE = (1(0.1) + 1(0.1) + 2(0.1) - 3)^2$
$= ((0.1 + 1.1 + 2.1) - 3)$
$= (3.3 - 3)^2$
$= (0.3)^2 = 0.09$

$J(\theta) = 0.09/2 = 0.045$

S2

$= (1(0.1) + 2(0.1) + 1(0.1) - 4)^2$
$= (3.3 - 4)^2 = (0.7)^2$
$= 0.49 = 0.245$

83

$$SSE = \left(1(0.1) + 3(0.1) + 3(0.1)\right) - 5)^2$$

$$= (6.3 - 8)^2$$

$$= (1.3)^2 = 1.69$$

$$= 0.845$$

87

$\alpha = 0.1$

$$\theta_0 = 0 - 0.1(1)$$

$$= +0.1$$

$$\theta_1 = 0 - 0.1(1)$$

$$= 0.1$$

$$\theta_2 = 0 - 0.1($$

$\alpha = 0.1$

$$\theta_0 = 0 + (0 - 3)(1)$$

$$= 0.1 \times 6$$

$$= \cancel{0.6}\ 0.3$$

$$\theta_1 = 0 + 0.1(9 - 3)(1)$$

$$= \cancel{0.1(6)}\ 0.3$$

$$= 0.6$$

$$\theta_2 = 0 + 0.1(9 - 3)(2)$$

$$= \cancel{0.1(12)}$$

$$= \cancel{1.2}\ 0.6$$

$$J(\theta) = 1.2\ \cancel{0.6}(1) + (0.6)(1) + (1.2) \times$$

2.8 → 3

0.2

$$= \cancel{0.6(1) + 0.6(1)} +$$

$$1.2(2) \quad 1.8 \to 3$$

$$= 0.6 + 0.6 + 2.4$$

$$= \cancel{2.4}\ \cancel{2.6}\ \cancel{0.6}\ 4.4$$

2

$$\cancel{SSE = 0.31}$$

89:

$$= (\cancel{2.4}\ 3.6 - 4)^2 = \cancel{0.16}$$

1.5

$\cancel{1.8}$ - 4

$$= \cancel{(0.36)}$$

$$= \cancel{0.16}$$

$$= \cancel{0.36}$$

1.44

2.5

$\cancel{2.89}$

h(x)

82!

$$\cancel{0.6(1)} +$$

$$= \cancel{0.36}$$

6.25

83!

$$0.6(1) + 0.6(3) + 0.2(3)$$

$$= 2.4 + 3.6$$

$$= (6 - 5) \quad = 9 - 5 = 2$$

n-dim vector

$$= \frac{\cancel{0.72}}{2}$$

4

$$= \boxed{\cancel{3.36}}$$

4+

$$1.52$$

$$J(\theta) = 0.76$$

① Closed form Solutions

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_d x_d$$

$$h_\theta(x) = \theta^T x$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}_{(d+1 \times 1)} \quad x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}_{(d+1 \times 1)}$$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{n} (h_\theta x^{(i)} - y^{(i)})^2$$

$$X = \begin{bmatrix} \dots x^{(1)T} \\ \dots x^{(2)T} \dots \\ \dots x^{(n)T} \dots \end{bmatrix}_{n \times d+1}$$

Design matrix

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}_{(n \times 1)} \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}_{(d+1 \times 1)}$$

if z is a n-dim vector.

then,

$$\sum_{i=1}^{n} z_i^2 = z^T z^{\text{vector}}$$

$$X\theta - \vec{y} = \begin{bmatrix} x^{(1)T}\theta - \vec{y} \\ x^{(2)T}\theta - \vec{y} \\ \vdots \\ x^{(n)T}\theta - \vec{y} \end{bmatrix}$$

X design matrix

$$\frac{1}{2}(X\theta - \vec{y})^T (X\theta - \vec{y})$$

$$= \frac{1}{2} \sum_{i=1}^{n} (h_\theta(x^{(i)}) - y^{(i)})^2$$

x_1	x_2	y
1	2	3
2	1	4
3	3	5

$x_0 = 1$

$$h_\theta x = \theta_0 x_1 + \theta_1 x_1 + \theta_2 x_2$$

$$\theta_0 = \theta_1 = \theta_2 = 0$$

S1 $\hat{y} = 0(1) + 0(1) + 2(0)$
$= 0$

$$SSE = (0-3)^2 + (0-4)^2 + (0-5)^2$$
$$= 9 + 16 + 25$$
$$= 50$$

$\boxed{J(\theta) = 25}$ $\quad \alpha = 0.1$

$$\theta_0 = 0 + (0-3)(1) \times (0.1)$$
$$= 0.3$$
$$\theta_1 = 0 + (3-0) \times (1) \times (0.1)$$
$$= 0.3$$
$$\theta_2 = 0 + 0.1(3-0)(2)$$
$$= 0.6$$

S1 $h_\theta x = 0.3(1) + 0.3(1) + 0.6(2)$
$= 1.8$

S2 $h_\theta x = 0.3(1) + 0.3(2) + 0.6(1)$
$= 1.5$

S3 $h_\theta x = 0.3(1) + 0.3(3) + 0.6(3)$
$= 3$

$$SSE = (3-1.8)^2 + (4-1.5)^2 + (5-3)^2$$
$$= (1.2)^2 + (2.5)^2 + 4$$

$= 1.44 + 6.25 + 4$

$SSE = 11.69$

$\boxed{J(\theta) = 5.845}$

$\theta_0 = 0.3 + (0.1)(1.2)(1)$

$= 0.3 + 0.12$

$= 0.42$

$\theta_1 = 0.3 + (0.1)(1.2)(1)$

$= 0.42$

$\theta_2 = 0.6 + (0.1)(1.2)(2)$

$= 0.84$

S1 $h_\theta x = 0.22(1) + 0.22(1) + 0.34(2)$

$= 1.12$

S2 $h_\theta x = 0.22(1) + 0.22(2) + 0.34(1)$

$= 1$

S3 $h_\theta x = 0.22(1) + 0.22(3) + 0.34(3)$

$= 1.9$

$SSE = (3 -$

S1 $h_\theta x = 0.42(1) + 0.42(1) + 0.84(2)$

$= 2.52$

S2 $h_\theta x = 0.42(1) + 0.42(2) + 0.84(1)$

$= 2.1$

S3 $h_\theta x = 0.42(1) + 0.42(3) + 0.84(3)$

$= 4.2$

$SSE = (3-2.52)^2 + (4-2.1)^2 + (5-4.2)^2$

$= (0.48)^2 + (1.9)^2 + (0.8)^2$

$= (0.2304) + 3.61 + 0.64$

$= 4.4804$

$\boxed{J(\theta) = 2.2402}$

$\theta_0 = 0.42 + 0.1(0.48)(1)$

$= 0.468$

$\theta_1 = 0.42 + 0.1(0.48)(1)$

$= 0.468$

$\theta_2 = 0.84 + 0.1(0.48)(2)$

$= 0.936$

S1 $h_\theta x = 0.468(1) + 0.468(1) + 0.936(2)$

$= 2.808$

S2 $h_\theta x = 0.468(1) + 0.468(2) + 0.936(1)$

$= 2.34$

S3 $h_\theta x = 0.468(1) + 0.468(3) + 0.936(3)$

$= 4.68$

$SSE = (3-2.808)^2 + (4-2.34)^2 + (5-4.68)^2$

$= (0.192)^2 + (1.66)^2 + (0.32)^2$

$= 0.036864 + 2.7556 + 0.1024$

$= 2.894864$

$\boxed{J(\theta) = 1.447432}$

L7 16/01/25

i) Calculate $\hat{y}$

ii) Error

$e = \hat{y} - y$

iii) Calculate gradient

$$\frac{\partial J}{\partial \theta_0} = \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_0^{(i)}$$

$= -3\times1 + -4\times1 + -5\times1$

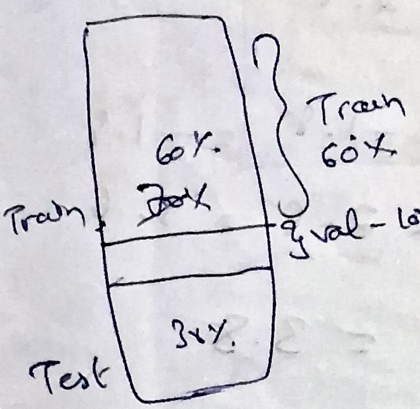
$= -12$

$\theta_0 = 1.2$

$\theta_1 = 2.6$

$\theta_2 = 2.5$

Data Splits:-



- validation set (val set) / Development set / Dev set

10 model

for each model perform 10-fold CV & obtain acc value for a val set.

Model	Acc → obtain by validation data.
LgR 1	83%
SVM 2	81%
Poly 6	79%
XGBoost	84%

std

20%

more robust

24/01/25 L15

3 fold CV

Acc std.

	x_1	x_2	y
#1	2	-1	1
#2	5	1.2	0
#3	1	2	1
#4	-3	-2	1
#5	4	0.1	0

Tr

F1. Fold : {1.8, 2.8}

F2 - {2.1, 3.1}

F3 - {1.9, 4}

5×7.9 / Lof

3.5

Fold - 1 :-

Train set:

#1	2	-1	1
#2	5	1.2	0
#3	1	2	1

Test 1

#4	-3	-2	1
#5	4	0.1	0

Fold 2:

Train set!

2	-1	1	#1
5	1.2	0	#2
-3	-2	1	#4
4	0.1	0	#5

Test set

1	2	1	#3

Fold 3'

Train set

1	2	1	#3
-3	-2	1	#4
4	0.1	0	#5

Test 1

2	-1	1	#1
5	1.2	0	#2

F1 :

$$g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

$$= \frac{1}{1+e^{-1.8}}$$

-1.8×2
$+ 2.8 \times 1$
$= -3.6$
$+ 2.8$

$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$

$\begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$

Tr Te

1, 2, 3, 4 5

θ −1.8, 2.8

$-1.8 \times 2 = -3.6$

$-1 \times 2.8 = -2.8$

$= -6.4$

$\cancel{-1.8 \times}$ $\cancel{2.8 \times}$ $\frac{1}{1+e^{6.4}}$

$= \frac{1}{\cancel{1+601.68}\ 1+601.84504}$

S1

$= \underline{0.001658}$

S2

5 1.2

$-1.8 \times 5 = -0.9$

$+2.8 \times 1.2 = -3.36$

$= \cancel{2.1}$

$= 2.46$

$\underline{0.9212}$

S3

-1.8×2.8

$1 \times -1.8 = -1.8$

$2 \times 2.8 = 5.6$

$= 3.8$

$= \cancel{0.02188}$

$= 0.978$

S4

−3, −2

$-3 \times -1.8 = 5.4$

$-2 \times 2.8 = -5.6$

$e^{0.2}$

$= 0.45016$

Error

$1 - 0.0016$

S1 $= 0.9984$

S2 $= -0.9212$

S3 $= 0.022$

S4 $= 0.54984$

P1 4, 0.1, 0

$4 \times -1.8 = -7.2$

$0.1 \times 2.8 = 0.28$

$= -6.92$

$e^{6.92}$

$= 1012.31$

$= 0.00098$

$= -0.00098$

$= 9.6 \times 10^{-7}$

100%

P3

#1, #2

1,	2	1
-3	-2	1

2.1
3.1

$1 \times 2.1 + 2 \times 3.1$

$= 2.1 + 6.2$

$= 8.3$ $e^{-8.3}$

$= 0.99$

1 $-3 \times 2.1 + 3.1 \times -2$

$= -6.3 + -6.2$

$= -12.5$

$= e^{12.5}$

$= 0.0060037$ 0

50%

P3

1.9, 4

1	2	1
-3	-2	1

$1 \times 1.9 = 1.9$

$2 \times 4 = 8$

$= 9.9$

$e^{-9.9} \rightarrow 0.999$

$-3 \times 1.9 = -5.7$

$-2 \times 4 = -8$

$= -13.7$

50%

$e^{13.7}$

$= 0.00000012$

---

n = no. of samples

k == n

LOOCV.

Leave One Out Cross-Validation

→ Data size less than 100

Train a Model:

1) Cross-validation
- k-fold
- LOOCV

2) Train - Test split

05/02/25

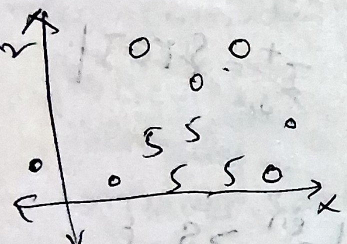
L21

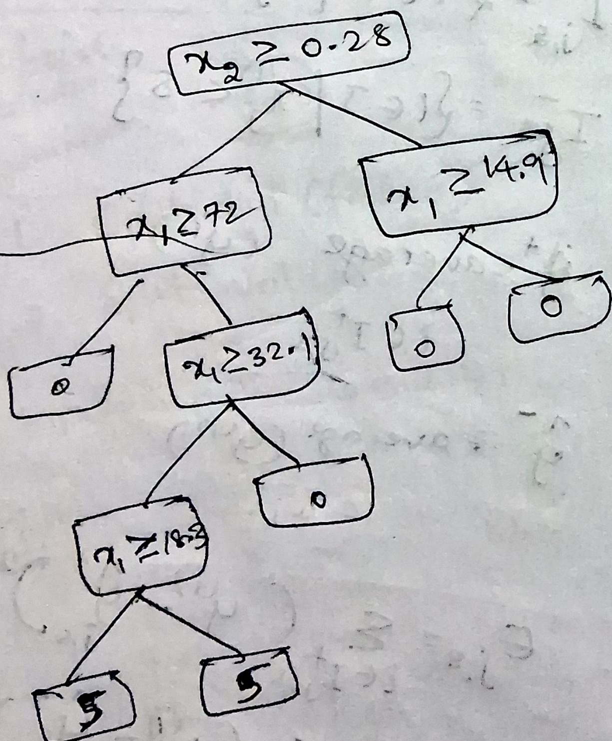
Set $j^*, s^* = \arg\min_{j,s} E_{j,s}$

return Node(j^*, s^*, BuildTree($I^+_{j^*s^*}$, k),
BuildTree($I^-_{j^*s^*}$, k))

}

Ex:

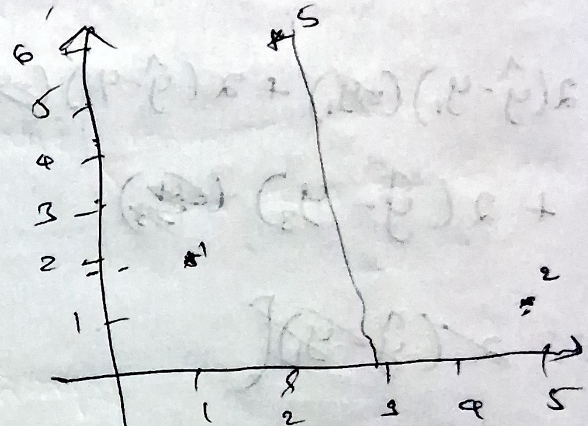




06/02/25

L22

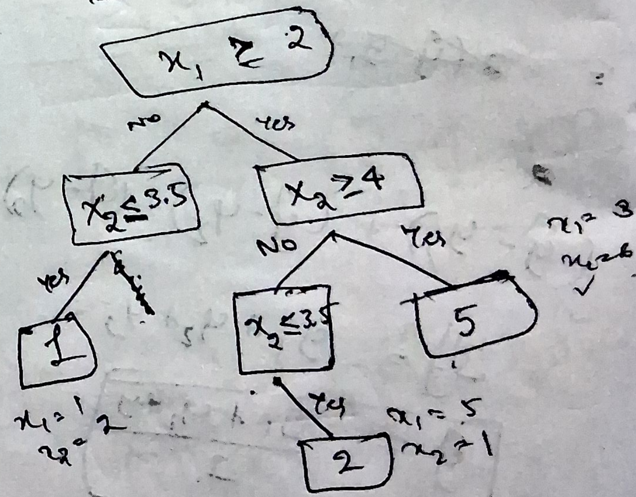
#	x_1	x_2	y
1	1	2	1
2	3	6	5
3	5	1	2



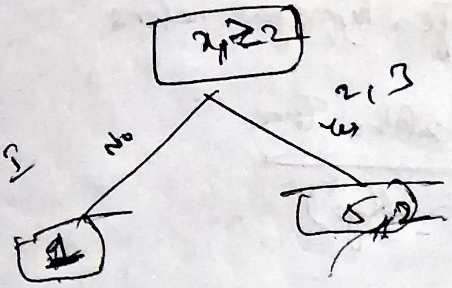
x_1: $s = \{2, 4\}$

x_2: $x_2 = \{4, 3.5\}$

for x_1



Tree 1
~~Step 1~~



$E^+ = \hat{y}^+ = \frac{5+2}{2} = 3.5$

$E^-_{x_1 \geq 2} = \hat{y}^- = \frac{1}{1} = 1$

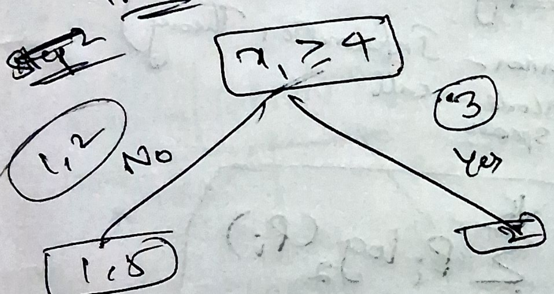
$E^- = 0$

$E^+_{x_1 \geq 2} = \left(0 + (5-3.5)^2 + (2-3.5)^2\right)$

$E = 4.5$

$\boxed{E = 4.5}$

Tree 2
~~Step 2~~



$E^+ = \hat{y}^+ = 2 = 2$

$E^-_{x_1 \geq 4} = \hat{y}^- = \frac{1+5}{2} = 3$

$E^+ = 0.$

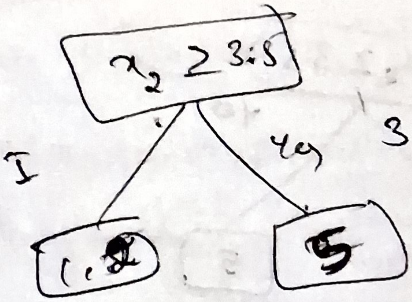
$E^- = (1-3)^2 + (5-3)^2$

$= 2^2 + 2^2$

$= 4 + 4 = 8$

$E = 8$

Tree 3
~~Step 3~~



$\hat{y}^- = \frac{1+2}{2} = 1.5$

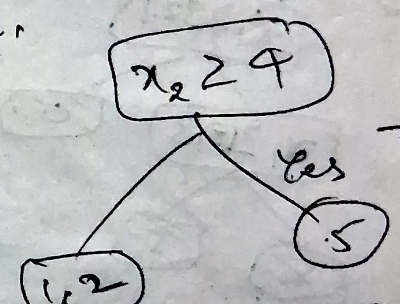
$y^+ = 5 = 0$

$E^+ = 0$

$E^- = \frac{1+2}{2} = 1.5$

$E = 0.5$

Tree 4



$\hat{y}^- = \frac{1+2}{2} = 1.5$

$\hat{y}^+ = 5$

$E^+ = 0$

$E^- = (1-1.5)^2 + (2-1.5)^2$

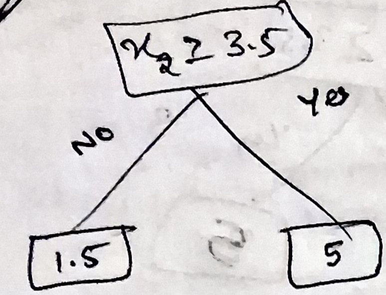
$E = 0.5$

$\min\{4.5, 8, 8, 0.5\}$

$g^*, s^* = x_2$

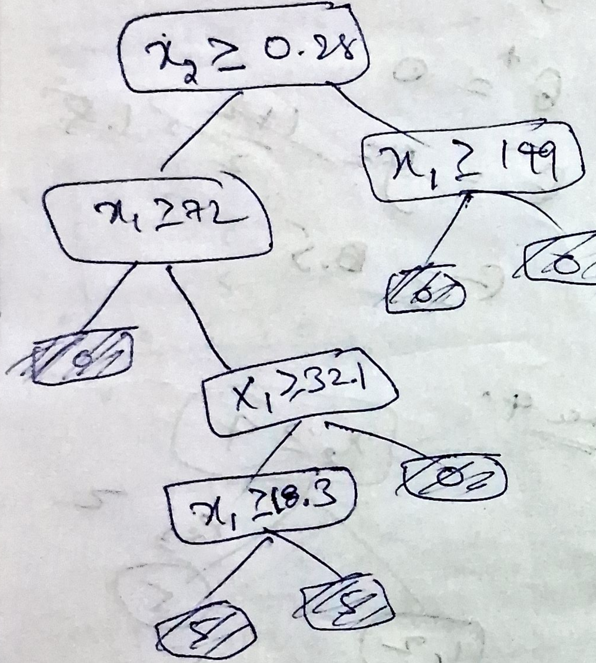
$s = 4$

Tree



07/02/25

L23



i) $C_\alpha(T) = \sum_{i=1}^{n} L(T(x^{(i)}, y^{(i)})$

$+ \alpha |T|$ → no. of leaves

~~Tree~~ grow / shrink

ii) Pruning

* DFT - Traversal
  - Start from leaf

* Grid search & minimize α of hyperparameters

Classification Tree

(CART Algo)

Nando DF UBC 2012

* Node - categorical value
  - more than 2 children

Entropy:

(Shannon - Information Theory)
encode speed → call

$$\boxed{-\sum_{i=1}^{K} P_i \log_2(P_i)}$$

IG = Information Gain

class $p=6$, class $n=6$ } 12 samples

$$H\left(\frac{p}{p+n}, \frac{n}{p+n}\right)$$

$$= -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\left(\frac{p}{p+n}\right)$$

$$= \sum_{i=1}^{K} -P_i \log_2 P_i$$

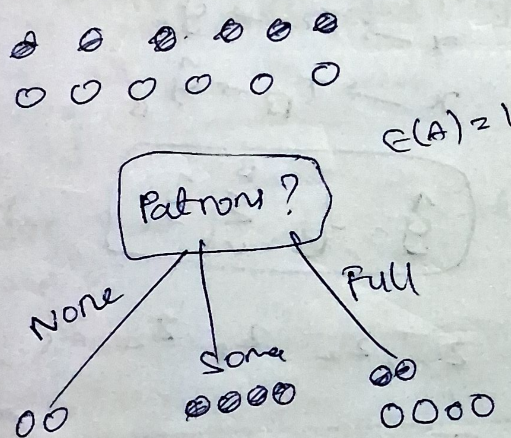
$$H\left(\frac{6}{12}, \frac{6}{12}\right)$$

$$= -\frac{6}{12}\log_2\left(\frac{1}{2}\right) - \frac{6}{12}\log_2\left(\frac{1}{2}\right)$$

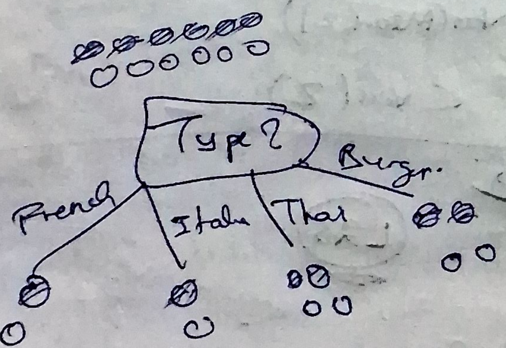
$$= -\frac{1}{2}(-1) - \left(\frac{1}{2}\right)(-1)$$

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

Case 1 (1)



Case 2 (2)



$$\boxed{EH(A) = \sum_{i=1}^{k} -\left(\frac{p_i+n_i}{p+n}\right) \cdot H\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)}$$

$$\boxed{IG(A) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - EH(A)}$$

maximum – Information Gain ✓

Entropy – Low ✓ – when homogenous

11/02/25

L24

Case 1

$$H(\text{patrons}) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right)$$

$$= H\left(\frac{6}{12}, \frac{6}{12}\right)$$

$$= 1$$

$$EH(\text{Patrons}) = \sum_{i=1}^{3} -\frac{p_i+n_i}{p+n} \cdot H\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$= -\left[\frac{p_1+n_1}{12} \cdot H\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{p_2+n_2}{12} H\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{p_3+n_3}{12} H\left(\frac{2}{6}, \frac{4}{6}\right)\right]$$

$$= -\left[\frac{2}{12} H(0,1) + \frac{4}{12} H(1,0) + \frac{6}{12} H\left(\frac{1}{3}, \frac{2}{3}\right)\right]$$

$$= -\left[\frac{2}{12}\left(-0\log_2 0 - 1\cdot\log_2 1\right) + \frac{4}{12}\left(1\log_2 1 - 0\log_2 0\right) + \frac{6}{12}\left(-\frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{4}{6}\log_2\left(\frac{4}{6}\right)\right)\right]$$

Random_forest(data)
{
df = data of size n.

B = no. of bags
new_data = [[]]*B
for b=1 to B
{
Draw a dataset $\tilde{D}_n$ by sampling with replacement.
new_data.append(x)
}

for data in new_data
{
Build_Tree(I, k, data)
Train a model $f^{(b)}$ on $\tilde{D}$
}
}

Build_Tree(I, k, ~~data~~ df)
{ $d = \sqrt{df.size.column}$
If (|I| ≤ k)
{
set $\hat{y}$ = average($y^{(i)}$)
return leaf(label = $\hat{y}$)
}
else
{
new_featurelist = random.sample(df.colums, d).

for each split dim f in new_feature_list
{
for each split value s
{
$I^{+}_{j,s}$
$I^{-}_{j,s}$

Precondition: A training set $S := (x_1, y_1) \ldots (x_n, y_n)$, feature F, & no. of trees in forest B.

function RandomForest(S, F)
H ← 0
for i ∈ 1, ..., B do

Random Forest

Age	salary	Spend pattern	Buy product
16	0	H	Y
30	90,000	L	N
31	90,000	H	N
85	80,000	M	N.

B = 2

b_1, b_2

B_1

Age	salary	spend pattern	Buy product
30	90,000	L	N
31	90,000	H	N
3~~0~~	90,000	H	N
~~56~~	1,00,000	~~H~~	N

B_2

16	0	H	Y
16	0	H	Y
31	99,000	H	N
16	0	H	Y

T_1

1. for each feature in (Age, salary).
{
for each value in feature.
{
IG = IG of value.append().
}

Boosting:-
- base learner/ weak learner:- small tree
- high bias.
- stump (1 node - Tree).

Inferences

$\hat{y} = T_1 + \lambda T_2 + \lambda T_3$

14/02/25

L27

Boosting
- weak learn
- Sequential learning, learn from previous tree mistakes.

* Bagging - Parallel
full blown decision tree

* Boosting - Sequential
(small weak learner
No Bootstrapping

Regression:

x_1	x_2	y
1	2	8.1
2	8	-1.5
3	4	3.8

1.) Initialize $r^{(i)} = y^{(i)}$ (residual)
$\forall\ i \in N$, where
N = no. of samples

$F(x, \theta) = 0$

2) for iteration $b = 1$ to B (no. of weak learner)
{
a) fit a model $f_b(x;\theta)$ to inputs x, output r.
b) update the model by adding shrunken version.

$F(x;\theta) = F(x,\theta) + \lambda f_b(x;\theta)$

c) update residuals

$r^{(i)} = r^{(i)} - \lambda f_b(x;\theta)$

}

}

Final Model.

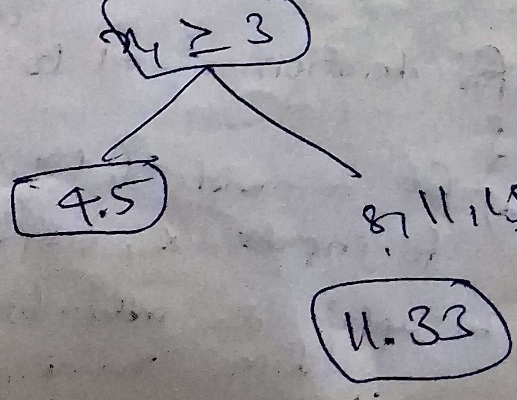
$$F(x;\theta) = \sum_{i=1}^{B} \lambda f_i(x;\theta)$$

x_1	y	r
1	3	3
2	6	6
3	8	8
4	11	4
5	15	15

step 1:

$r = y$

a) fit a model.



b) r =

x_1	y	$\hat{y}$	
1	3	~~4.5~~	3 + 0.1(4.5) = 3.45
2	6	~~4.5~~	6.45
3	8	~~11.3~~	9.13
4	11	~~11.3~~	12.13
5	15	~~11.3~~	16.13

$\lambda = 0.1$

c) r_{new}

-0.45

-0.45

-1.13

-1.13

-1.13

step 2:

$x_1 \geq 2.1$

www.kamperh.com/data414/notes/17_ensemble_methods_notes.pdf

18/2/25

L28

1. Boosted Regression Tree (AdaBoost Regression).

Boosted Classification Tree (AdaBoost Classification).

1). ~~Initialize $f(x)=a$~~

~~$r=y$~~

Ex:

x_1	x_2	y
1	2	4
2	3	5
3	4	6
4	5	7

$F(x) = 0$

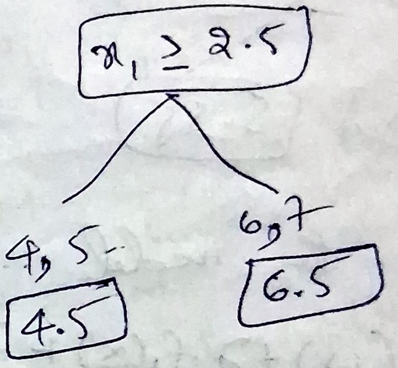
$f_0(x) = 0$

Iter 1:-

$r = y$

x_1	x_2	r
1	2	4
2	3	5
3	4	6
4	5	7

a) fit a model, $f_1(x) \rightarrow r$



b) update the model

$\lambda = 0.1$

$F(x) = f_0(x) + \lambda f_1(x)$

c). update residuals

$r^{(1)} = r^{(1)} - \lambda f_1(x^{(1)}) = 4 - 4.5 (0.1)$
$= 3.55$

$r^{(2)} = r^{(2)} - \lambda f_1(x^{(2)}) = 5 - 0.45$
$= 4.55$

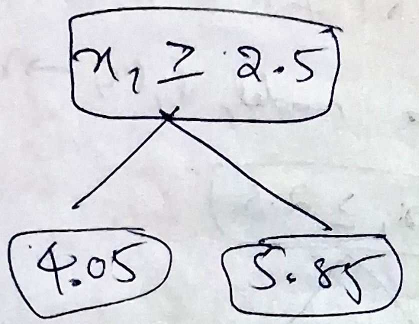
$r^{3} = r^{(3)} - \lambda f_1(x^{(3)}) = 6 - 0.65$
$= 5.35$

$r^{4} = r^{(4)} - \lambda f_1(x^{(4)}) = 7 - 0.63$
$= 6.35$

Iter 2:-

$$r_2 \begin{bmatrix}3.55\\4.55\\5.35\\6.35\end{bmatrix}$$

→ fit a model



b) update the model

$$F = f_0(x) + \lambda f_1(x) + \lambda f_2(x)$$

c) update residuals

$$r^{(1)} = 3.55 - 0.405 = 3.145$$
$$r^{(2)} = 4.55 - 0.405 = 4.145$$
$$r^{(3)} = 5.35 - 0.585 = 4.765$$
$$r^{(4)} = 6.35 - 0.585 = 5.765$$

$$F(x,\theta) = f_0(x) + \lambda f_1(x) + \lambda f_2(x) + \dots \lambda f_B(x)$$

Adaboost classification:-

Setup:-

- Binary classification
- Multiple weak learner models with target $\in \{-1, +1\}$
- combine models

$$F(x) = \text{sign}\left[\sum_{b=1}^{B} \lambda_b f_b(x, \theta_b)\right]$$

① Initialize training sample weights $w^{(n)} = 1$ for all samples

② for iteration $b = 1$ to B.
{

a) Fit a model $f_b(x, \theta_b)$ so that it minimizes classification error weighted by $w^{(n)}$.

b) Set model weight using error, ϵ

$$\lambda_b = \frac{1}{2}\log\left(\frac{1-\epsilon}{\epsilon}\right)$$

c) update training sample weights.

$$w^{(n)} = w^{n} e^{-\lambda_b},$$

if $f_b(x^{(n)}, \theta_b)$ is correct classification.