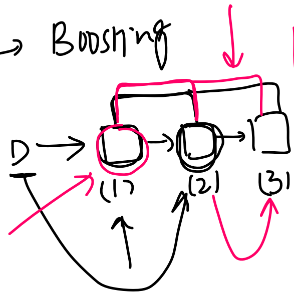
Gradient Boosting

04 August 2023 20:07

Ensemble technique

↳ Boosting

stagewise



↘ Additive modelling

weak learner

+ HB/LV

$\boxed{B \propto \frac{1}{V}}$

$\underbrace{HB + HB + HB}_{LB}$

LBLV

Regression

↓

Classification

Classification Vs Regression

04 August 2023 20:14

Gradient Boosting

mse / log loss

Input: training set $\{(x_i, y_i)\}_{i=1}^{n}$, a differentiable loss function $\boxed{L(y, F(x))}$, number of iterations M.

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$.
2. For $m = 1$ to M:
   (a) For $i = 1, 2, \ldots, N$ compute

   $$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}.$$

   (b) Fit a regression tree to the targets r_{im} giving terminal regions $R_{jm},\ j = 1, 2, \ldots, J_m$.

   (c) For $j = 1, 2, \ldots, J_m$ compute

   $$\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L\left(y_i, f_{m-1}(x_i) + \gamma\right).$$

   (d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.
3. Output $\hat{f}(x) = f_M(x)$.

Example

04 August 2023 20:15

8 rows 3 cols

$\downarrow \left(\frac{5}{3}\right)$

$\boxed{3} + \square + \square$ — $f_0(x)$, $f_1(x)$, $f_2(x)$ — mean, simple mean, decision tree, simple, DT, error

Pars

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

log odds

3 stages

pseudo residual

$y_i - f_0(x_i)$

reg → mean

class → log odds

cgpa	iq	is_placed
6.82	118	0
6.36	125	1
5.39	99	1
5.50	106	1
6.39	148	0
9.13	148	1
7.17	147	1
7.72	72	0

$\boxed{5 \; 1}$ $\boxed{3 \; 0}$

Stage 1 → simple model

↳ log(odds) → $\log\left(\frac{\#1}{\#0}\right)$

$\boxed{\log\left(\frac{5}{3}\right)} = f_0(x)$ 1st model

↳ $\boxed{0.51}$ → $\boxed{\log_e}$ $\log_{10} x$

$p = \frac{1}{1+e^{-\log odds}} = \frac{1}{1+e^{-0.51}} = 0.62$

prob, initial pred

cgpa	iq	is_placed	pre1(log-odds)
6.82	118	0	0.510826
6.36	125	1	0.510826
5.39	99	1	0.510826
5.50	106	1	0.510826
6.39	148	0	0.510826
9.13	148	1	0.510826
7.17	147	1	0.510826
7.72	72	0	0.510826

prob → $\frac{e^x}{1+e^x}$

y_i p_i

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)
6.82	118	0	0.510826	0.625
6.36	125	1	0.510826	0.625
5.39	99	1	0.510826	0.625
5.50	106	1	0.510826	0.625
6.39	148	0	0.510826	0.625
9.13	148	1	0.510826	0.625
7.17	147	1	0.510826	0.625
7.72	72	0	0.510826	0.625

$\boxed{0.62}$ → good starting point

$\boxed{\boxed{1} + \boxed{2}}$ → $f_1(x)$, dt, $f_2(x)$ → prediction

$\boxed{\text{log odds}}$

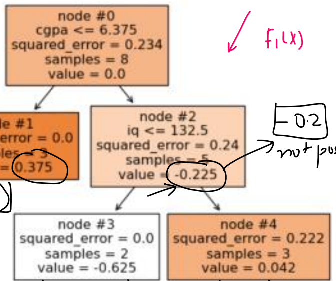
res1

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1
6.82	118	0	0.510826	0.625	-0.625
6.36	125	1	0.510826	0.625	0.375
5.39	99	1	0.510826	0.625	0.375
5.50	106	1	0.510826	0.625	0.375
6.39	148	0	0.510826	0.625	-0.625
9.13	148	1	0.510826	0.625	0.375
7.17	147	1	0.510826	0.625	0.375
7.72	72	0	0.510826	0.625	-0.625

→ number → diff of prob

dt — regression tree — weak learner → leaf nodes = 3

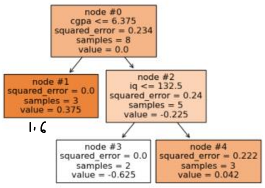
$f_1(x)$



1.6 — $-\infty$ $+\infty$ — -0.2 not possible — -2.66 — 0.18

0.51 — $\square + \square$ dt

$\boxed{F_0(X) \; 0.51}$ +



1.6

$$\frac{\sum Residual}{\sum [PreviousProb * (1 - PreviousProb)]}$$

$= \frac{-1}{0.375} = -2.66$

$\frac{-0.625 - 0.625}{0.625(1-0.625) + 0.625(1-0.625)}$

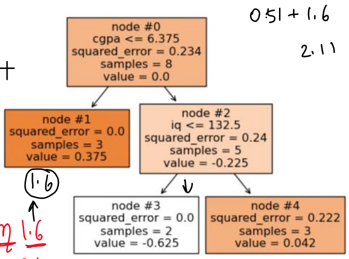
$\frac{-2 \times 0.625}{0.625 \times 0.375 + 0.625 \times 0.375} = \frac{-\cancel{2 \times 0.625}}{\cancel{2 \times 0.625} \times 0.375}$

output → log odds

$0.51 + (-2.66) = -2.15$

$0.51 + 1.6$ — 2.11

$0.51 +$



learning rate

1.6 — $0.51 + \eta \, 1.6$ — 0.1 — 0.16

-2.66 — 0.18 — -0.6 -0.1 residual

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1	leaf_entry1
6.82	118	0	0.510826	0.625	-0.625	3
6.36	125	1	0.510826	0.625	0.375	1
5.39	99	1	0.510826	0.625	0.375	1
5.50	106	1	0.510826	0.625	0.375	1
6.39	148	0	0.510826	0.625	-0.625	4
9.13	148	1	0.510826	0.625	0.375	4
7.17	147	1	0.510826	0.625	0.375	4
7.72	72	0	0.510826	0.625	-0.625	3

pre2(log-odds)
-2.159174
2.110826
2.110826
2.110826
0.690826
0.690826
0.690826
-2.159174

$p = \frac{1}{1+e^{-\log odds}}$

$\frac{1}{1+e^{+2.159}} = 0.10$

$\boxed{\square + \square}$ → $\boxed{0.10}$ → 50% 0.5 → $\boxed{0}$

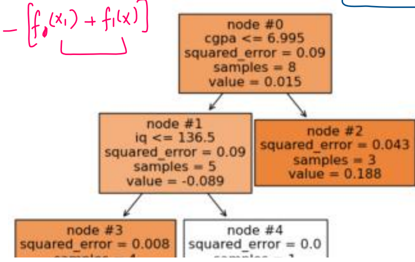
$\boxed{\square + \square}$ → log odds → prob → residuals

$\boxed{\square + \square + \square}$ dt1 dt2 ↳ log odds

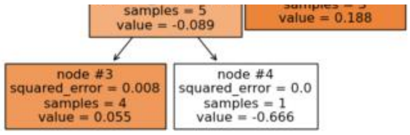
$y - [f_0(x_i) + f_1(x)]$

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1	leaf_entry1	pre2(log-odds)	pre2(probability)
6.82	118	0	0.510826	0.625	-0.625	3	-2.159174	0.103477
6.36	125	1	0.510826	0.625	0.375	1	2.110826	0.891951
5.39	99	1	0.510826	0.625	0.375	1	2.110826	0.891951
5.50	106	1	0.510826	0.625	0.375	1	2.110826	0.891951
6.39	148	0	0.510826	0.625	-0.625	4	0.690826	0.666151
9.13	148	1	0.510826	0.625	0.375	4	0.690826	0.666151
7.17	147	1	0.510826	0.625	0.375	4	0.690826	0.666151

res2
-0.103477
0.108049
0.108049
0.108049
-0.666151
0.333849
0.333849



6.39	148	0	0.510826	0.625	-0.625	4	0.690826	0.666151	-0.666151
9.13	148	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849
7.17	147	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849
7.72	72	0	0.510826	0.625	-0.625	3	-2.159174	0.103477	-0.103477

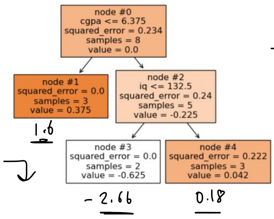


$F_1(x)$ $\qquad$ $F_2(x)$

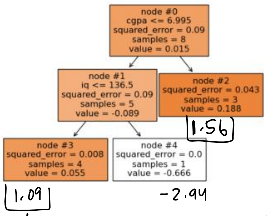
$F_0(X)$

0.51

+



+



$F_0(x) + F_1(x)$

output

↓

log odds

$$\frac{-0.66}{\cancel{0.66}\times(1-0.66)} = \frac{-1}{0.34}$$

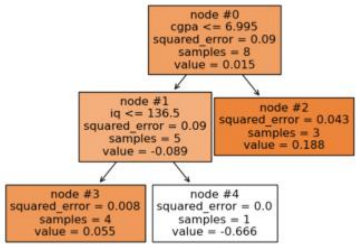
$$\frac{\sum Residual}{\sum[PreviousProb*(1-PreviousProb)]}$$

$-2.15 + 1.09$

$= -1.06$

$2.11 + 1.09 = 3.20$

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1	leaf_entry1	pre2(log-odds)	pre2(probability)	res2	leaf_entry2
6.82	118	0	0.510826	0.625	-0.625	3	-2.159174	0.103477	-0.103477	3
6.36	125	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3
5.39	99	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3
5.50	106	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3
6.39	148	0	0.510826	0.625	-0.625	4	0.690826	0.666151	-0.666151	4
9.13	148	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849	2
7.17	147	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849	2
7.72	72	0	0.510826	0.625	-0.625	3	-2.159174	0.103477	-0.103477	2



$$\frac{1}{1+e^{+1.06}}$$

out → $p = \frac{1}{1+e^{-\log odds}}$ 0

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1	leaf_entry1	pre2(log-odds)	pre2(probability)	res2	leaf_entry2	pre3(log-odds)	pre3(probability)
6.82	118	0	0.510826	0.625	-0.625	3	-2.159174	0.103477	-0.103477	3	-1.068349	0.255717
6.36	125	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3	3.201651	0.960896
5.39	99	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3	3.201651	0.960896
5.50	106	1	0.510826	0.625	0.375	1	2.110826	0.891951	0.108049	3	3.201651	0.960896
6.39	148	0	0.510826	0.625	-0.625	4	0.690826	0.666151	-0.666151	4	-1.798349	0.142052
9.13	148	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849	2	2.251651	0.904793
7.17	147	1	0.510826	0.625	0.375	4	0.690826	0.666151	0.333849	2	2.251651	0.904793
7.72	72	0	0.510826	0.625	-0.625	3	-2.159174	0.103477	-0.103477	2	-0.598349	0.354722

~~f_1~~ f $\qquad$ $\boxed{X_1 \quad X_2}$ $\quad$ y $\quad$ pred $\quad$ $\boxed{res\ 1}$

↓

dt 1

purple → $F_0(x)$

red + yellow → $F_0(x) + F_1(x)$ → purple

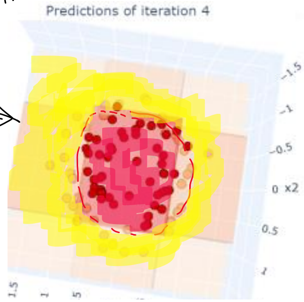
↳ dt

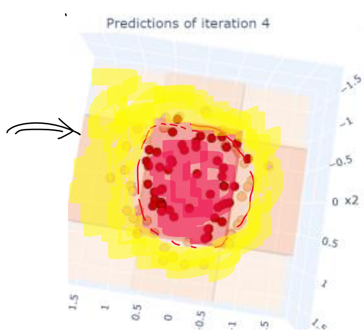
$F_0(x) + F_1(x) + F_2(x)$

↓ $\qquad$ ↓

dt $\qquad$ dt

Predictions of iteration 4



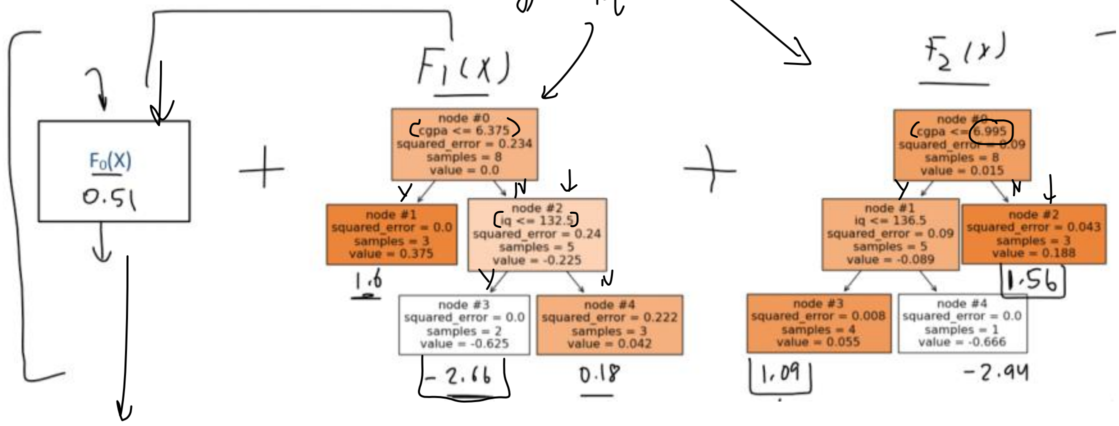

Predictions of iteration 4
x2

Prediction

07 August 2023 04:53

$\{7.2, 100\} \rightarrow \{0, 1\} \rightarrow \boxed{0}$

cgpa iq



xq - log odds = -0.59 → prob → output

0.35

0.5

$\frac{1}{1+e^{0.59}}$

$0.51 + (-2.66) + 1.56 = \boxed{-0.59}$

Geometric Intuition

04 August 2023 20:15

Maths behind Classification

08 August 2023 22:54

$f_0(x)$ $f_1(x)$ $f_2(x)$.

Input: training set $\{(x_i, y_i)\}_{i=1}^{n}$, a differentiable loss function $L(y, F(x))$, number of iterations M.

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$. min $\log\left(\frac{P_{avs}}{1-P_{avs}}\right)$
2. For $m = 1$ to M:
   (a) For $i = 1, 2, \ldots, N$ compute

   $i \to$ row $\quad r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}$. $\quad \frac{\partial L}{\partial f_0(x_i)}$

   (b) Fit a regression tree to the targets r_{im} giving terminal regions $R_{jm},\ j = 1, 2, \ldots, J_m$.

   (c) For $j = 1, 2, \ldots, J_m$ compute

   $$\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L\left(y_i, f_{m-1}(x_i) + \gamma\right).$$

   $$\frac{\sum Residual}{\sum[PreviousProb * (1 - PreviousProb)]}$$

   (d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.
3. Output $\hat{f}(x) = f_M(x)$.

Step 0 -> The Loss Function

10 August 2023 09:43

Log loss $\rightarrow$ $L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log p_i + (1-y_i)\log(1-p_i) \rightarrow L(p_i)$

$\downarrow$

$L(\log(\text{odds}))$

$\hat{y}_i \rightarrow$ output prob

cgpa	iq	placement	$\hat{y}(p_i)$
8	80	1	0.62
7	70	0	0.32
6	60	1	0.51

$\boxed{\text{log odds}} \rightarrow \boxed{\text{prob}}$

log odds

$$\boxed{\log\left(\frac{p_i}{1-p_i}\right)}$$

$$L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log p_i + (1-y_i)\log(1-p_i)$$

$$L = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log p_i + \log(1-p_i) - y_i \log(1-p_i)\right]$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i\left[\log p_i - \log(1-p_i)\right] + \log(1-p_i)\right]$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(\text{odd}_i) + \log(1-p_i)\right]$$

$$\boxed{p_i = \frac{e^{\log(\text{odd}_i)}}{1+e^{\log(\text{odd}_i)}}}$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(\text{odd}_i) + \log\left(1 - \frac{e^{\log(\text{odd}_i)}}{1+e^{\log(\text{odd}_i)}}\right)\right]$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(\text{odd}_i) + \log\left(\frac{1}{1+e^{\log(\text{odds}_i)}}\right)\right]$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(\text{odds}_i) + \log 1 - \log\left(1+e^{\log(\text{odds}_i)}\right)\right]$$

$$L = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(\text{odd}_i) - \log\left(1+e^{\log(\text{odd}_i)}\right)\right] \leftarrow \text{data}$$

$$L = -\frac{1}{n}\left[\sum_{i=1} y_i \log(odds) - \log(\ldots)\right]$$

$$\boxed{L = -y \log(odds) + \log\left(1 + e^{\underline{\log(odds)}}\right)}$$

Step 1 -> Finding $F_0(x)$

10 August 2023 09:44

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} \underbrace{L(y_i, \gamma)}_{\text{Loss function}}$.

$$L = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \log(odds_i) - \log(1+e^{\log(odds_i)})\right]$$

$L(y_i, \log(odd_i))$

$\hookrightarrow L(y_i, \gamma)$

$$\underset{\gamma}{\arg\min}\ L = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i \gamma - \log(1+e^{\gamma})\right]$$

cgpa | iq | placement

1
0
1
0

$$\frac{1+1+0+0}{4}$$

$$\boxed{\frac{2}{4}}$$

p(1)

Pavg

$$\frac{\partial L}{\partial \gamma} = -\frac{1}{n}\left[\sum_{i=1}^{n} y_i - \frac{e^{\gamma}}{1+e^{\gamma}}\right] = 0$$

$$= \underbrace{-\frac{1}{n}\sum_{i=1}^{n} y_i}_{} + \frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{e^{\gamma}}{1+e^{\gamma}}}_{} = 0$$

$$= -P_{avg} + \frac{\cancel{n}}{\cancel{n}}\frac{e^{\gamma}}{1+e^{\gamma}} = 0$$

$$\frac{e^{\gamma}}{1+e^{\gamma}} = P_{avg}$$

$$e^{\gamma} = P_{avg} + P_{avg}e^{\gamma} \Rightarrow e^{\gamma} - P_{avg}e^{\gamma} = P_{avg}$$

$$e^{\gamma}(1-P_{avg}) = P_{avg}$$

$$e^{\gamma} = \frac{P_{avg}}{1-P_{avg}}$$

$$\boxed{\gamma = \log\left(\frac{P_{avg}}{1-P_{avg}}\right)}$$

$$f_0(x) = \gamma = \log\left(\frac{P_{avg}}{1-P_{avg}}\right)$$

$$\log\left(\frac{5/8}{1-5/8}\right) = \log\left(\frac{5/\cancel{8}}{3/\cancel{8}}\right) = \log\left(\frac{5}{3}\right)$$

Step 2.a -> Pseudo Residuals

10 August 2023 11:23

n rows

2. For $m = 1$ to M:

   (a) For $i = 1, 2, \ldots, N$ compute

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}.$$

$f(x) \rightarrow f_0(x) \rightarrow$ log odds

$$L = -y \log(\text{odds}) + \log(1+e^{\log(\text{odds})})$$

$$\frac{\partial L}{\partial \log(\text{odds})} = -y + \boxed{\frac{e^{\log(\text{odds})}}{1+e^{\log(\text{odds})}}}$$

$\rightarrow$ prob output

$$= -y + p = \boxed{y - p} \rightarrow \boxed{y_i - p_i}$$

log odds

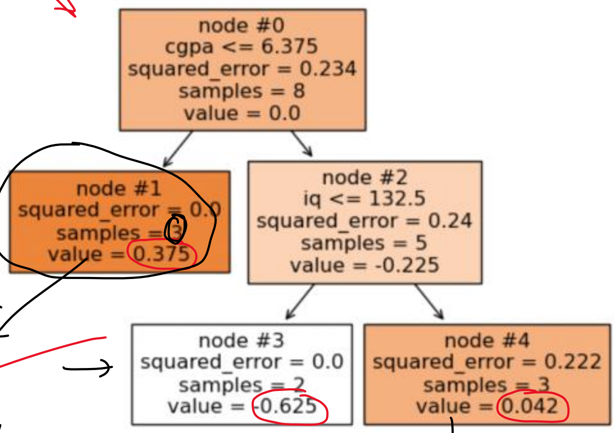
Step 2.b -> Train Regression Tree

10 August 2023 11:41

X f0(x)

cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1
6.82	118	0	0.510826	0.625	-0.625
6.36	125	1	0.510826	0.625	0.375
5.39	99	1	0.510826	0.625	0.375
5.50	106	1	0.510826	0.625	0.375
6.39	148	0	0.510826	0.625	-0.625
9.13	148	1	0.510826	0.625	0.375
7.17	147	1	0.510826	0.625	0.375
7.72	72	0	0.510826	0.625	-0.625

regression



output

from leaf node

$$\left\{ \frac{\sum Residual}{\sum[PreviousProb * (1 - PreviousProb)]} \right\}$$

Step 2.c -> Compute Lambda for all leaf nodes

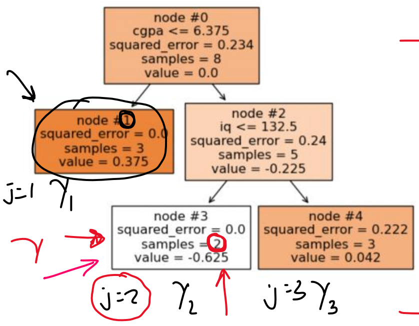
10 August 2023 12:02

$m = 1$

(c) For $j = 1, 2, \ldots, J_m$ compute

γ_1

$$\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{jm}} L\left(y_i, f_{m-1}(x_i) + \gamma\right).$$



cgpa	iq	is_placed	pre1(log-odds)	pre1(probability)	res1	leaf_entry1
6.82	118	0	0.510826	0.625	-0.625	3
6.36	125	1	0.510826	0.625	0.375	1
5.39	99	1	0.510826	0.625	0.375	1
5.50	106	1	0.510826	0.625	0.375	1
6.39	148	0	0.510826	0.625	-0.625	4
9.13	148	1	0.510826	0.625	0.375	4
7.17	147	1	0.510826	0.625	0.375	4
7.72	72	0	0.510826	0.625	-0.625	3

$$\gamma_2 = \arg\min_{\gamma} \; L(y_1, f_0(x_1) + \gamma) + L(y_8, f_0(x_8) + \gamma) \qquad L'$$

minimize

$$\frac{res1}{p_1(1-p_1)} + \frac{res8}{p_8(1-p_8)}$$

$$\frac{\partial L'}{\partial \gamma} = 0$$

$x \to f_0(x_1) + \gamma$

$a \to f_0(x_1)$

Taylor series (upto 2nd derivative)

$$\gamma_2 = \frac{\sum \text{residual}_i}{\sum \text{prob}_i(1-\text{prob}_i)}$$

$$\underset{x \to a}{f(x)} = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots$$

$$L = L(y_1, f_0(x_1)) + \frac{\partial L}{\partial f_0(x_1)}\gamma + \frac{1}{2}\frac{\partial^2 L}{\partial f_0(x_1)^2}\gamma^2$$

$$\frac{\partial L}{\partial \gamma_1} = \frac{\partial L}{\partial f_0(x_1)} + \frac{\cancel{2}}{\cancel{2}}\gamma\frac{\partial^2 L}{\partial f_0(x_1)^2} = 0$$

$$\gamma \frac{\partial^2 L}{\partial f_0(x_1)^2} = -\frac{\partial L}{\partial f_0(x_1)}$$

$$\gamma = \frac{-\frac{\partial L}{\partial f_0(x_1)}}{\frac{\partial^2 L}{\partial f_0(x_1)^2}}$$

$\to$ residual wrt row1 $(y_1 - p_1)$

$p_1(1-p_1)$

$$\frac{\partial L}{\partial f_0(x_1)} = -\left(-y_1 + \frac{e^{\log(odds_1)}}{1+e^{\log(odds_1)}}\right) \to p_1 \quad = \boxed{y_1 - p_1} \quad \text{residual}$$

$$\boxed{\frac{\partial L}{\partial f_0(x_1)}} = -\left[-y_1 + \boxed{\frac{e^{\log(odds_1)}}{1+e^{\log(odds_1)}}}\right]$$

$$\frac{\partial}{\partial f_0(x_1)} \frac{\partial L}{\partial f_0(x_1)} = \frac{\partial}{\partial \log(odds_1)}\left(y_1 - \frac{e^{\log(odds_1)}}{1+e^{\log(odds_1)}}\right)$$

$$\frac{\partial}{\partial \log(odd)}\left[y_1 - \underbrace{e^{\log(odd_1)}}_{x}\ \underbrace{(1+e^{\log(odd_1)})^{-1}}_{y}\right]$$

$$x'y + xy'$$

$$= -\left[e^{\log(odd_1)}(1+e^{\log(odds_1)})^{-1} - e^{\log(odd_1)}(1+e^{\log(odds_1)})^{-2}\right]$$

$$= \frac{e^{\log(odds_1)}}{(1+e^{\log(odds_1)})^2} - \frac{e^{\log(odds_1)}}{(1+e^{\log(odds_1)})}$$

$$= \frac{e^{\log(odds_1)}}{1+e^{\log(odd_1)}}\left[\frac{1}{1+e^{\log(odd_1)}} - 1\right] \rightarrow \frac{1}{1+e^{\log(odds_1)}}$$

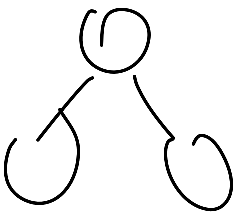
$$= \underbrace{\boxed{\frac{e^{\log(odds_1)}}{(1+e^{\log(odds_1)})}}}_{P_1}\ \underbrace{\boxed{\frac{1}{(1+e^{\log(odds_1)})}}}_{(1-P_1)}$$

Step 2d - Update the model

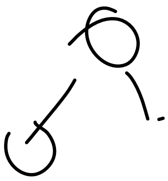
10 August 2023 12:10

$$f_1(x) = f_0(x) + \text{output from dt} \left\{ \text{some leaf node} \right\}$$

$f_0(x) \downarrow$ logodds



$f_2(x)$



$f_3(x) \longrightarrow$ M dcis

Step 3 - Final Model

10 August 2023 12:10

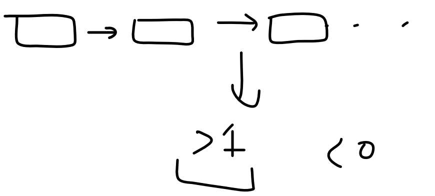
$$f_M(x) = f_{m-1}(x) + \text{last decision output}$$

Boosting model

Log(odds) Vs Probability

10 August 2023 13:06

$-\infty$ $+\infty$ $(0 - 1)$



① Unconstrained Prediction Space: Log odds can span the entire real line (—∞ to +∞), while probabilities are constrained between 0 and 1. Algorithms like gradient boosting involve adding corrections (via the weak learners) to the predictions iteratively. If you're working in the log odds space, there's no need to worry about your predictions going out of bounds.

$-\infty$ - $+\infty$

10 0.9 0.5 0.25

② Better Gradients: When computing gradients (which guide the addition of new trees in boosting), the gradients can be more informative and have better magnitudes in the log odds space than in the probability space, especially when probabilities are near 0 or 1.

1 0 0.1

0.1 0.05 0

$-\infty$ ∞