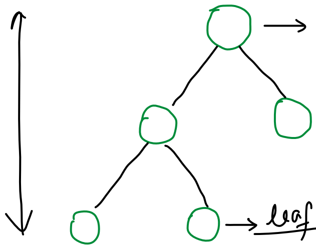
Plan of Attack

30 November 2023 10:42



Regression

$$\text{Similarity Score} = \frac{\left(\sum \text{Residuals}\right)^2}{N + \lambda} \quad ①$$

Classification

$$\text{Similarity Score} = \frac{\left(\sum \text{Residuals}\right)^2}{\sum^{N}[P(1-P)] + \lambda} \quad ②$$

reg

$$\text{Output value} = \frac{\sum \text{Residuals}}{N + \lambda} \quad ③$$

$$\text{Output Value} = \frac{\left(\sum \text{Residuals}\right)}{\sum^{N}[P(1-P)] + \lambda} \quad ④$$

$\boxed{1}$ intro

$\boxed{2}$ regression

$\boxed{3}$ classification

deep dive → goal

Prerequisite

30 November 2023 10:44

→ Xgboost → Reg
↘ Classi

→ Gradient Boosting →

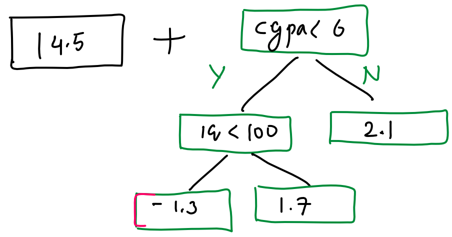
→ Tree (basic)

Boosting as an [Additive Model]

30 November 2023 12:04

Xgboost → 2 model

cgpa	iq	package
6.5	71	13
5	91	10
9.1	120	16



Xgboost

train

out leaf weight

{9, 100}

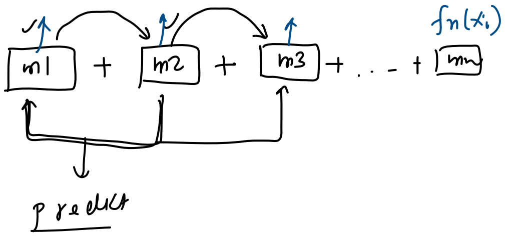
$\hat{y}_i = 14.5 + 2.1$

$= \boxed{16.6}$ lpa

$\hat{y}_i = f_1(x_i) + f_2(x_i) + \dots + f_n(x_i)$

$$\boxed{\hat{y}_i = \sum_{i=1}^{n} f_i(x_i)} \rightarrow \text{xgboost}$$

$\boxed{m_1 + m_2}$



[reg]

$$\boxed{\text{output}} = \boxed{\frac{\text{sum of residuals}}{N + \lambda}}$$

$1 + \lambda$

origin story

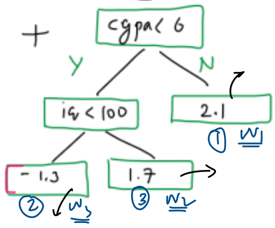
XGBoost Loss Function

30 November 2023 13:19

cgpa	iq	package $\boxed{y_i}$
6.5	71	13
5	91	10
9.1	120	16

m1 $\boxed{14.5}$ + m2

5 leaf



$\hat{y}_i = m_1 + m_2$

$f_1(x_i) + \boxed{f_2(x_i)}$ → value

14.5

3 output values

→ loss function ?

$\boxed{\hat{y}_i} = \sum_{i=1}^{2} f_i(x_i)$

reg mse

$L(y_i, \hat{y}_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

$$\left[\text{output} = \frac{\sum \text{sum of residual}}{N + \lambda}\right]$$

number of residuals

$\boxed{y_i \quad \hat{y}_i}$ $\boxed{\text{reduce}}$

$y_i \quad \hat{y}_i$ → relation

↳ loss function → minimize

$\boxed{\text{Xgboost}}$ → loss functions

$y_i \rightarrow \hat{y}_i \rightarrow$ relation

Gradient Boosting –

flexible → loss function
↳ differentiable

differentiable → loss-function → regularization parameter

$[L(y_i, \hat{y}_i)]$

mse   log loss

$m_1 + \boxed{m_2}$ → dt

base

optimization

$$\boxed{L = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \boxed{\Omega(f_2(x_i))}}$$ → minimize

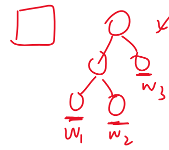
$L = \sum_{i=1}^{n} L(y_i, \hat{y}_i)$ loss

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

3

regularization   # leaf nodes   regulators

t



$\frac{1}{2}\lambda(w_1^2 + w_2^2 + w_3^2)$

The Derivation

30 November 2023 13:57

objective func

$$\mathcal{L} = \underbrace{\sum_{i=1}^{n} L(y_i, \hat{y}_i)}_{\text{loss function}} + \underbrace{\Omega(f_k)}_{\text{regular}}$$

$\gamma T + \frac{1}{2}\lambda \|w\|^2$

T → leaf

$\|w\|^2$ → leaf w

boosting → mean → ☐ ☐ ☐

xgboost → 1 model ☐ → objective

Stage I

mean

m1

$f_1(x_i)$ → $\mathcal{L}^{\langle 1 \rangle} = \sum_{i=1}^{n} L(y_i, \hat{y}_i)$

f_2 f_1 X

Stage II

mean

m1 $f_1(x_i)$

m2 $f_2(x_i)$ → decision

$$\mathcal{L}^{\langle 2 \rangle} = \sum_{i=1}^{n} L(y_i, f_1(x_i) + f_2(x_i)) + \Omega(f_2(x_i))$$

$\hat{y}_i = f_1(x_i) + f_2(x_i)$

Stage III

mean

m1 $f_1(x_i)$

m2 $f_2(x_i)$

m3 $f_3(x_i)$

$$\mathcal{L}^{\langle 3 \rangle} = \sum_{i=1}^{n} (y_i, f_1(x_i) + f_2(x_i) + f_3(x_i)) + \Omega(f_3(x_i))$$

$f_1(x_i) + f_2(x_i) + \dots f_{t-1}(x_i) \rightarrow \hat{y}_i^{\langle t-1 \rangle}$

Stage $\boxed{t}$

mean

m1 $f_1(x_i)$

m2 $f_2(x_i)$

. . . .

mt $f_t(x_i)$

last decis

$$\mathcal{L}^{\langle t \rangle} = \sum_{i=1}^{n} L(y_i, \underbrace{f_1(x_i) + f_2(x_i) + \dots + f_t(x_i)}) + \Omega(f_t(x_i))$$

$$\mathcal{L}^{\langle t \rangle} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{\langle t-1 \rangle} + f_t(x_i)) + \Omega(f_t)$$

At any stage t

$$\underbrace{\mathcal{L}^{\langle t \rangle}} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{\langle t-1 \rangle} + f_t(x_i)) + \Omega(f_t(x_i))$$

$f_t(x_i)$

→ minimize

Problem with Objective Function

30 November 2023 20:40

$$\mathcal{L}^{<t>} = \boxed{\sum_{i=1}^{n} L(y_i, \hat{y}_i^{<t-1>} + f_t(x_i))} + \Omega(f_t(x_i))$$

deis

optimize

$$\frac{\partial L}{\partial f_t(x_i)} = 0$$

diffrentiable

loss

minize

[optimization] ~~yes~~

$\hat{y}_i$

obj fn

approximate

inp | out (y_i)

exp | salary

x_i

$$\hat{y}_i = m(exp_i) + b$$

Linear reg

$$L = \sum_{i=1}^{n} L(y_i, \hat{y}_i) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$L = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

minimum

$$\frac{\partial L}{\partial m} = 0 \qquad \frac{\partial L}{\partial b} = 0$$

$m \Rightarrow$

(m, b) → $\min_{b}$ grad desc

$L \rightarrow \mathcal{L}^{<t>}$

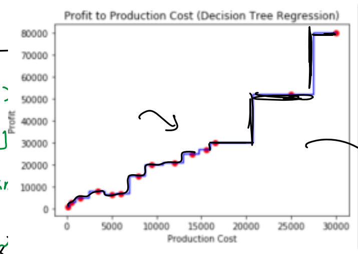
$m, b \quad f_t(x_i)$

$\hat{y}_i = mx + b$

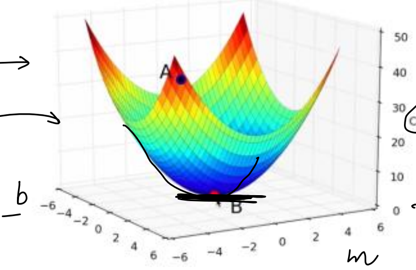
$\hat{y}_i = \underbrace{\hat{y}_i^{<t-1>}}_{\text{constant}}$

sal, linc, exp

sal, exp



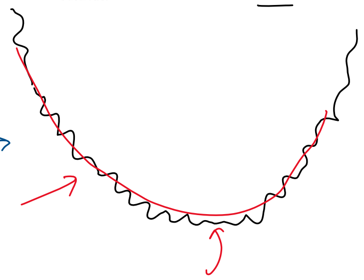
piecewise constant



$$\left[\frac{\partial L}{\partial m}, \frac{\partial L}{\partial b}\right] = 0$$

Cost / Loss

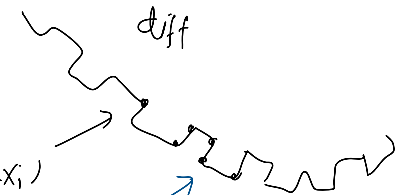
$L(\hat{y}_i)$



some

often → smooth

→ diffrentable



diff

$f_t(x_i)$

non-diff

$$\frac{\partial L}{\partial f_t(x_i)}$$

Taylor series

The Solution - Taylor Series

01 December 2023 00:02

powerful approx techqi

Objective → usable

↓

Taylor series

complex function → approx → polynomial

↓

$\underline{e^x}$ ⟶ $\underline{3x^2+2}$

accurate approx

↑

$3x^2 + 3x^3 + x^4 + x^5 \ldots$

How

$e^x \rightarrow$ polynomial

→ derivatives

$\boxed{e^x} \rightarrow e^x$

$f(x)$ point → $\boxed{a}$

$$f(x) \simeq f(a) + \underbrace{\frac{f'(a)}{1!}(x-a)} + \underbrace{\frac{f''(a)}{2!}(x-a)^2} + \underbrace{\frac{f'''(a)}{3!}(x-a)^3} + \ldots$$

$f(x) = e^x$ $\boxed{a=0}$

$f(a) = e^a = 1$

$$\frac{f'(a)}{1!}(x-a) = \frac{e^a}{1}(x-a) = e^0(x-0) = x$$

$$\frac{f''(a)}{2!}(x-a)^2 = \frac{e^a}{2!}(x-a)^2 = \frac{e^0(x)^2}{2!} = \frac{x^2}{2}$$

$$\frac{f'''(a)}{3!}(x-a)^3 = \frac{e^a}{3!}(x-a)^3 = \frac{1}{6}x^3 = \frac{x^3}{6}$$

$$e^x \simeq \boxed{1+x} + \frac{x^2}{2} + \frac{x^3}{6} + \ldots$$

Applying Taylor Series

01 December 2023 00:51

$$f(x) \simeq f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 \ldots$$

$$\frac{\partial f}{\partial x} =$$

$f(x) = x^2$

$f'(x)\ \frac{\partial f}{\partial x} = 2x$

$f'(a) = 2a$

$$\mathcal{L}^{<t>} = \left[\sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{<t-1>} + f_t(x_i)\right)\right] + \Omega(f_t(x_i)) \rightarrow 2^{nd} \text{ degree}$$

$x \rightarrow \hat{y}_i^{<t-1>} + f_t(x_i) \quad a \rightarrow \hat{y}_i^{<t-1>}$

$$\mathcal{L}^{<t>} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{<t-1>}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) + \Omega(f_t(x_i))$$

$$f(a) = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{<t-1>})$$

gradient g_i

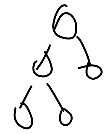
$$f'(a)(x-a) = \sum_{i=1}^{n} \boxed{\frac{\partial L(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>}}} f_t(x_i)$$

hessian h_i

$$\frac{f''(a)}{2}(x-a)^2 = \frac{1}{2}\sum_{i=1}^{n} \boxed{\frac{\partial^2 L(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>2}}} f_t^2(x_i)$$

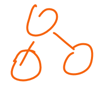
$$\mathcal{L}^{\langle t\rangle} \simeq \sum_{i=1}^{n}\Big[\underbrace{L(y_i,\hat{y}_i^{\langle t-1\rangle})}_{\to 0} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\Big] + \Omega(f_t(x_i))$$

$$\mathcal{L}^{\langle t\rangle} = \sum_{i=1}^{n}\Big[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\Big] + \boxed{\Omega(f_t(x_i))} \to \text{expand}$$



T leaf nodes

$$\mathcal{L}^{\langle t\rangle} = \underbrace{\sum_{i=1}^{n}\Big[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\Big]}_{i=1\to n \;\; \#\text{rows}} + \gamma T + \underbrace{\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2}_{j=1\to T,\;\; T\to \#\text{leafnodes}}$$



$T=2$

$j=1,2$

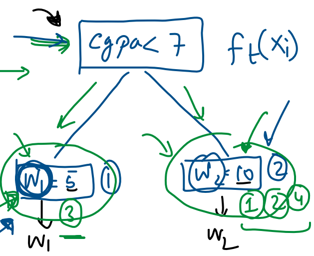
$w_j \to f_t(x_i)$

$\sum_{i=1}^{n} \to \sum_{j=1}^{T}\sum_{i\in I_j}$

$$\mathcal{L}^{\langle t\rangle} = \sum_{j=1}^{T}\Big[\sum_{i\in I_j} g_i w_j + \frac{1}{2}\sum_{i\in I_j} h_i w_j^2\Big] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

100

$$\mathcal{L}^{\langle t\rangle} = \sum_{i=1}^{n}\Big[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\Big] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$



	x_i cgpa	x_i package
1	7.1	4.5
2	8.2	6.9
3	6.4	5.1
4	9.1	10

100

$$\mathcal{L}^{\langle t\rangle} = \sum_{j=1}^{T}\Big[\sum_{i\in I_j} g_i w_j + \frac{1}{2}\sum_{i\in I_j} h_i w_j^2\Big] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

$\boxed{\sum_{i=1}^{n} \to \sum_{j=1}^{T}\sum_{i\in I_j}}$

$i\in I_j \to$ instance set of leaf j

2 times

$f_t(x_i) \to w_j$

4 terms

$f_t(x_3)$

decision

$$\to \Big(g_1 f_t(x_1) + \frac{1}{2} h_1 f_t^2(x_1)\Big) + \Big(g_2 f_t(x_2) + \frac{1}{2} h_2 f_t^2(x_2)\Big) + \dots + g_4 f_t(x_4) + \frac{1}{2} h_4 f_t^2(x_4)$$

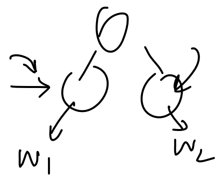
$j=1,\ j=2$

$$\to \underbrace{g_3 f_t(x_3) + \frac{1}{2} h_3 f_t^2(x_3)}_{g_3 w_1 + \frac{1}{2} h_3 w_1^2} + \underbrace{g_1 f_t(x_1) + \frac{1}{2} h_1 f_t^2(x_1)}_{g_1 w_2 + \frac{1}{2} h_1 w_2^2} + g_2 f_t(x_2) + \frac{1}{2} h_2 f_t^2(x_2) + g_4 f_t(x_4) + \frac{1}{2} h_4 f_t^2(x_4)$$

$$\mathcal{L}^{\langle t\rangle} = \sum_{j=1}^{T}\Big[\sum_{i\in I_j} g_i w_j + \frac{1}{2}\sum_{i\in I_j} h_i w_j^2\Big] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

$$\mathcal{L}^{\langle t\rangle} = \sum_{j=1}^{T}\Big[\sum_{i\in I_j} g_i w_j + \frac{1}{2}\sum_{i\in I_j} h_i w_j^2 + \frac{1}{2}\lambda w_j^2\Big] + \gamma T$$

$$\mathcal{L}^{\langle t\rangle} = \sum_{j=1}^{T}\Big[\sum_{i\in I_j} g_i w_j + \frac{1}{2}\Big(\sum_{i\in I_j} h_i + \lambda\Big) w_j^2\Big] + \gamma T$$



$$\mathcal{L}^{\langle t \rangle} = \sum_{j=1}^{T} \left[\underline{\sum_{i \in I_j} g_i w_j} + \frac{1}{2} \underline{\left(\sum_{i \in I_j} h_i + \lambda \right)} w_j^2 \right] + \gamma T$$

w_1 w_2

$\frac{\partial \mathcal{L}}{\partial w_j}$

$$\frac{\partial \mathcal{L}^{\langle t \rangle}}{\partial w_j} = \sum_{i \in I_j} g_i + \left(\sum_{i \in I_j} h_i + \lambda \right) w_j = 0$$

$$\boxed{w_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}}$$

$g_i / h_i \rightarrow$ formula the output

Output Value for Regression

01 December 2023 13:45

$$w_j = \frac{\boxed{-\sum_{i \in I_j} g_i}}{\sum_{i \in I_j} h_i + \lambda}$$

→ gradient

→ hessian

$$g = \frac{\partial L(y_i, \hat{y}_i^{\langle t-1 \rangle})}{\partial \hat{y}_i^{\langle t-1 \rangle}} \qquad h = \frac{\partial^2 L(y_i, \hat{y}_i^{\langle t-1 \rangle})}{\partial \hat{y}_i^{\langle t-1 \rangle 2}}$$

Loss → reg, classi

Reg → mse → $\boxed{L = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

prev step

$$g_i = \frac{\partial L}{\partial \hat{y}_i} = -\sum_{i=1}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(\hat{y}_i - y_i)$$

$$= \frac{\sum_{i \in I_j}(y_i - \hat{y}_i^{\langle t-1 \rangle})}{\sum_{i \in I_j} 1 + \lambda} \rightarrow \frac{\text{sum of residual}}{\text{\# of residuals} + \lambda}$$

→ residual for every point

cgpa	iq	package (y_i ✓)	pred ($\hat{y}_i^{t-1}$ ✓)	residual
8	80	8	8.5	8-8.5
9	90	9	8.5	9-8.5

$$\frac{\partial}{\partial \hat{y}_i}\frac{\partial L}{\partial \hat{y}_i} = \frac{\partial}{\partial \hat{y}_i}(\hat{y}_i - y_i) = 1$$

$$w_j = \frac{\text{sum of residual}}{\text{\# residuals} + \lambda}$$

Output Value for Classification

01 December 2023 13:45

$$w_j = \frac{-\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

$$g = \frac{\partial L(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>}}$$

$$h = \frac{\partial^2 L(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>2}}$$

loss - logloss → gradient boosting

$$\left[w_j = \frac{\text{sum of residual}}{\sum_{i \in I_j} P_i(1-p_i) + \lambda} \right]$$

P_i → pred prob of prev timestep

$$\boxed{L^{\langle t\rangle}} = \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right) w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i+\lambda\right) w_j^2\right] + \gamma T$$

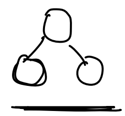
$$w_j = \frac{-\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i+\lambda}$$

$$L^{\langle t\rangle}(q) = \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right)\frac{\left(-\sum_{i\in I_j} g_i\right)}{\sum_{i\in I_j} h_i+\lambda} + \frac{1}{2}\cancel{\left(\sum_{i\in I_j} h_i+\lambda\right)}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\left(\sum_{i\in I_j} h_i+\lambda\right)^{\cancel{2}}}\right] + \gamma T$$

$$L^{\langle t\rangle}(q) = \sum_{j=1}^{T}\left[-\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i+\lambda} + \frac{1}{2}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i+\lambda}\right] + \gamma T$$

$$\text{Similarity score} = \sum_{j=1}^{T}\left[-\frac{1}{2}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i+\lambda}\right] + \gamma T$$

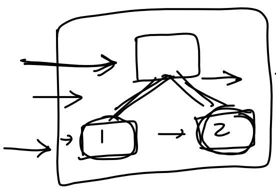
iter t



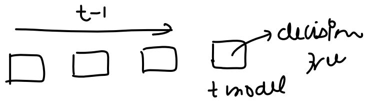
$$\left[\; L^{\langle t\rangle}(q) = -\frac{1}{2}\sum_{j=1}^{\overset{2\,\leftarrow}{T}}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i+\lambda} + \gamma T \;\right]$$

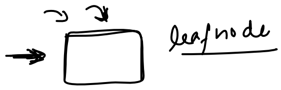
gini / entropy

1



$\xrightarrow{t-1}$





$$L^{\langle t\rangle}(q_p) = -\frac{1}{2}\frac{\left(\sum_{i\in I_p} g_i\right)^2}{\sum_{i\in I_p} h_i+\lambda} + \gamma$$

re

$$\boxed{L^{\langle t\rangle}(q_c)} = -\frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i+\lambda}\right] + 2\gamma$$

$$L^{\langle t\rangle}(q_p) - L^{\langle t\rangle}(q_c) \qquad -\gamma$$

$$-\frac{1}{}\left(\sum_{i\in I_p} g_i\right)^2 \quad \left(-\frac{1}{2}\left[\left(\sum_{i\in I_L} g_i\right)^2 \quad \left(\sum_{i\in I_R} g_i\right)^2\right] \pm 2\gamma\right)$$

$$-\frac{1}{2}\frac{\left(\sum_{i\in I_P} g_i\right)^2}{\sum_{i\in I_P} h_i+\lambda} + \gamma - \left(-\frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i+\lambda}\right] \pm 2\gamma\right)$$

$$\left[-\frac{1}{2}\frac{\left(\sum_{i\in I_P} g_i\right)^2}{\sum_{i\in I_P} h_i+\lambda} + \frac{1}{2}\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{1}{2}\frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i+\lambda}\right] - \gamma$$

$$\frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i+\lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i+\lambda} - \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_P} h_i+\lambda}\right] - \underline{\gamma}$$

differ

Final Calculation of Similarity

02 December 2023 01:27

$$\text{Similarity score} = \sum_{j=1}^{T} \left[-\frac{1}{2} \frac{\left(\sum_{i \in I_j} \underline{g_i}\right)^2}{\sum_{i \in I_j} \underline{h_i} + \lambda} \right] + \gamma T$$

reg

$g \rightarrow$ sum of residuals

$h \rightarrow \sum 1 \rightarrow N$

Class

$g \rightarrow$ residual

$h \rightarrow p_i(1-p_i)$