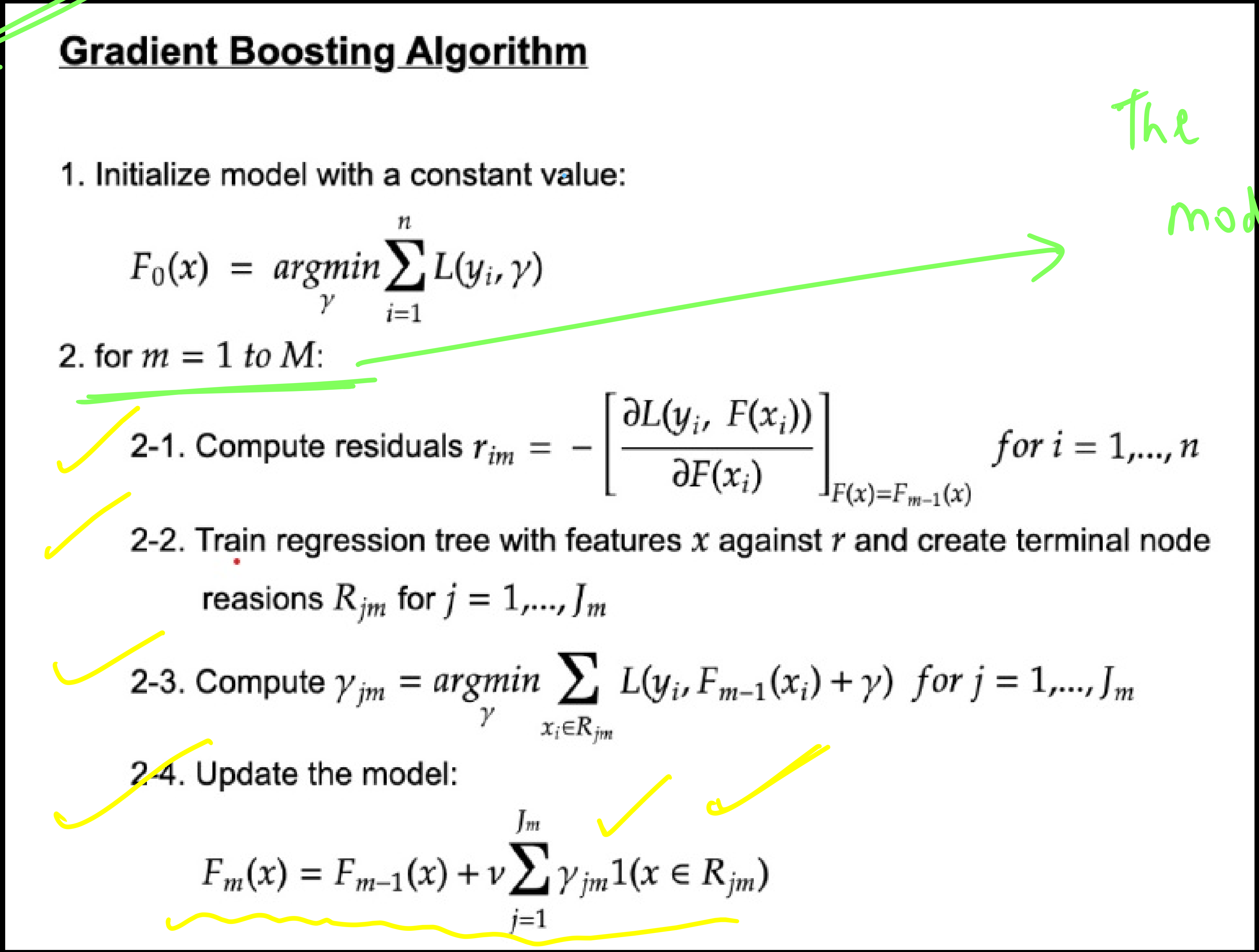
Agenda

- → Boosting (GBDT)
- → Bias Variance
- → Hyperparameter
- → Stochastic GBDT

Friedmann



Gradient Boosting Algorithm

1. Initialize model with a constant value:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. for $m = 1$ *to* M:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ *for* $i = 1,\ldots,n$

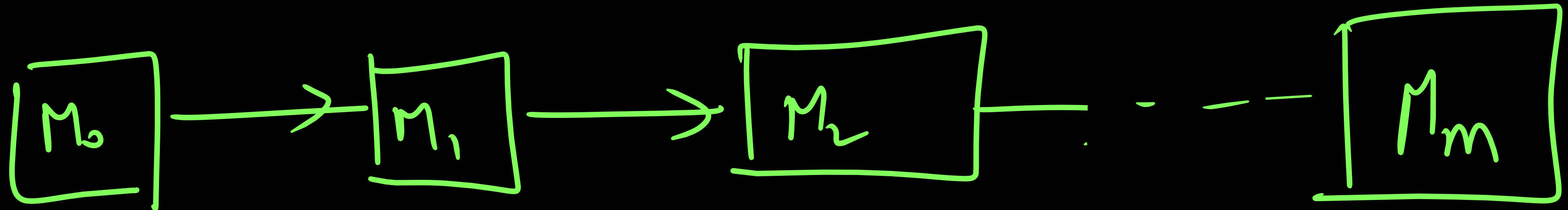
2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\ldots,J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ *for* $j = 1,\ldots,J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

The number of models we built



X

y

Height	Gender	Weight(y)
1.6	M	82
1.5	F	55
1.4	F	61
1.4	M	65

$f_0(x) = \hat{y}_0$	residual $= y - F_0(x)$
65.75	16.25
65.75	-10.75
65.75	-4.75
65.75	-0.75

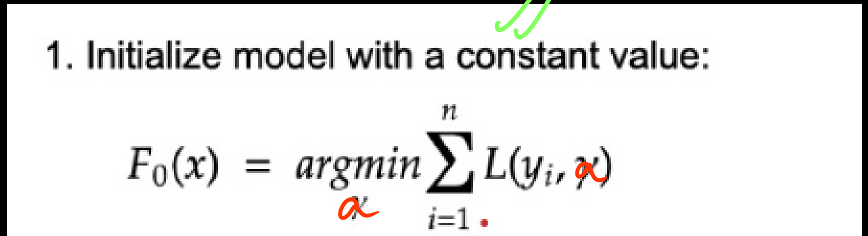
$\rightarrow M_0$

$$\hat{y}_0 = \frac{82 + 55 + 61 + 65}{4}$$

$$= 65.75$$

$$\boxed{65.75}$$

M_0



1. Initialize model with a constant value:

$$F_0(x) = \underset{\alpha}{argmin} \sum_{i=1}^{n} L(y_i, \overset{\alpha}{\gamma})$$

$$y = \begin{bmatrix} \overset{0}{2}, & \overset{1}{4}, & \overset{2}{1} \end{bmatrix}$$

$\rightarrow 1, 2$

$$= \underset{a}{argmin} \sum_{i=1}^{n} \left(y^{(i)} - a\right)^2$$

$$\rightarrow \frac{\partial L}{\partial a} = -2 \sum_{i=1}^{n} \left(y^{(i)} - a\right) = 0$$

$$\rightarrow \sum_{i=1}^{n} \left(y^{(i)} - a\right) = 0$$

$$\rightarrow \sum_{i=1}^{n} y^{(i)} - \sum_{i=1}^{n} a = 0$$

$$\rightarrow \sum_{i=1}^{n} y^{(i)} - na = 0$$

$$\rightarrow \boxed{\frac{\sum_{i=1}^{n} y^{(i)}}{n} = a}$$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\dots,J_m$

e_0

Height	Gender	Weight(y)	err$_0$
1.6	M	82	16.25
1.5	F	55	-10.75
1.4	F	61	-4.75
1.4	M	65	-0.75

0.1

1

$F_1 = F_0 + V\hat{y}_1$

$\hat{y}_1$	F_0	F_1	F_1	2
7.75	65.75	66.52	73.5	81.25
-7.75	65.75	64.97	58	50.25
-7.75	65.75	64.97	58	50.25
7.75	65.75	66.52	73.5	

R_{j1}

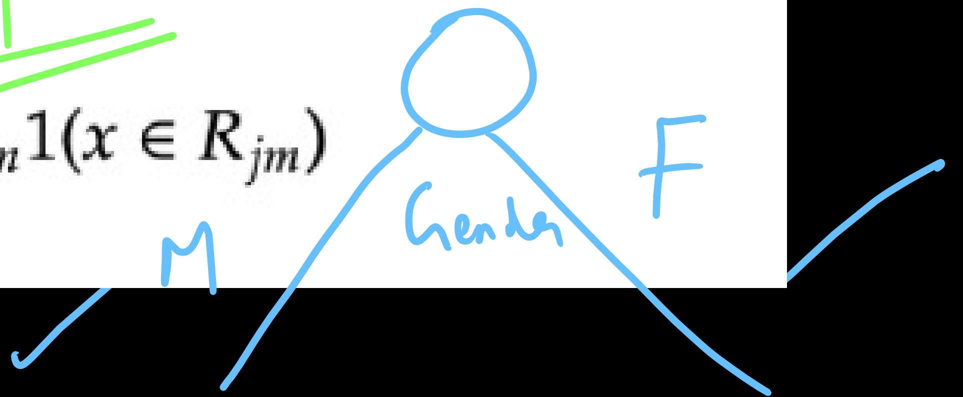
2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

M1

$\boxed{M_0}$ +

65.75

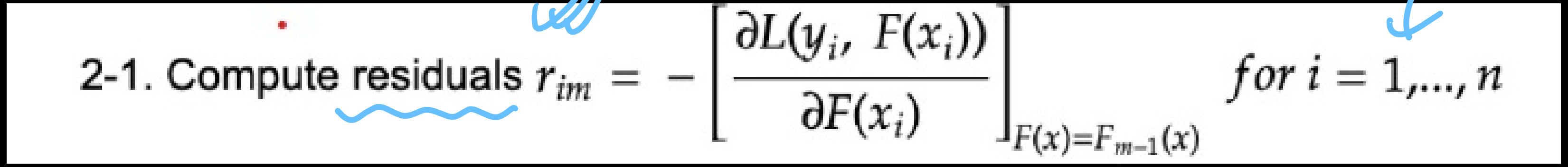


$R_{1,1}$ $\gamma_{1,1} = \frac{16.25 - 0.75}{2}$

$= 7.75$

$\gamma_{2,1} = \frac{-10.75 - 4.75}{2}$ $R_{2,1}$

$= -7.75$



2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i,\, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ for $i = 1,\dots,n$

↳ Why residual is $y^{(i)} - f_0(x^{(i)})$?

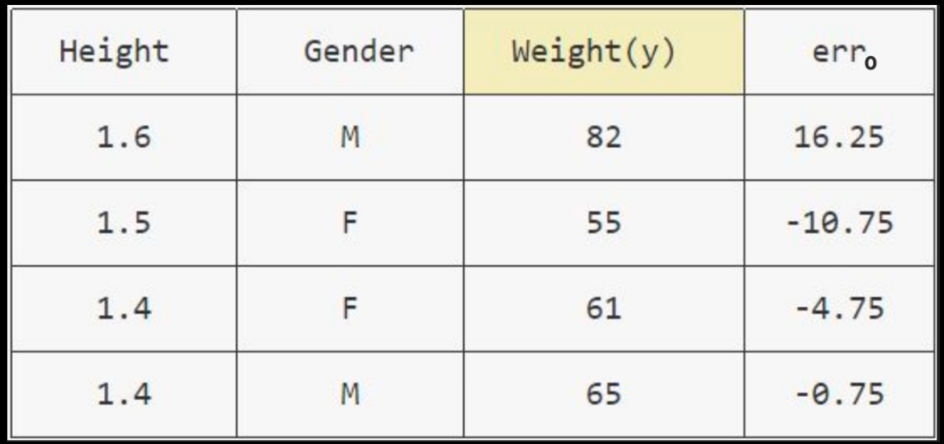
$$L(y^{(i)}\; F(x^{(i)})) = \sum (y^{(i)} - F(x^{(i)}))^2$$

$$\frac{\partial L}{\partial F(x^{(i)})} = \underset{\hookrightarrow \text{ Ignore } 2}{-2} \cdot (y^{(i)} - F(x^{(i)}))$$

$$r_{im} = -\frac{\partial L}{\partial F(x^{(i)})} = (y^{(i)} - F(x^{(i)}))\Big|_{F_{m-1}(x)}$$

$$= y^{(i)} - F_{m-1}(x^{(i)})$$

e_0



Height	Gender	Weight(y)	err_0
1.6	M	82	16.25
1.5	F	55	-10.75
1.4	F	61	-4.75
1.4	M	65	-0.75

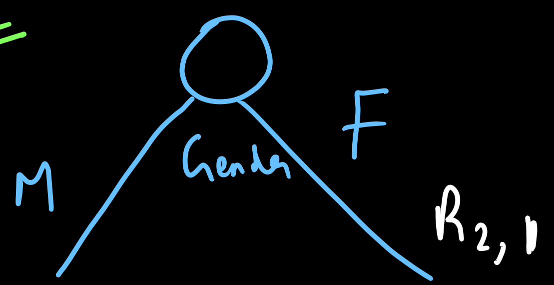
1.8 M

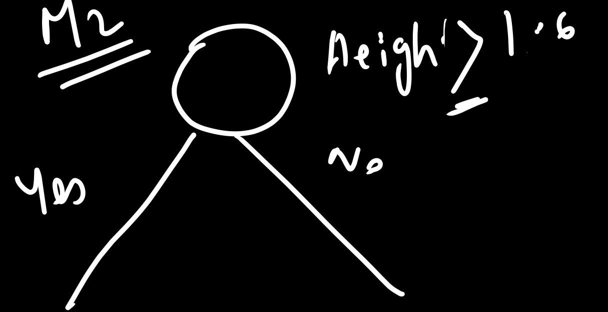
$F_1 = F_0 + \gamma \hat{y}_1$	y	e_1	$\hat{y}_2$	$F_2 = F_1 + \gamma \hat{y}_2$
66.52	82	15.48	15.48	68.06
64.97	55	-9.97	-5.15	64.455
64.97	61	-3.97	-5.15	64.455
66.52	65	-1.52	-5.15	66

$\boxed{M_0}$ + M1 + M2

65.75

$R_{1,1}$





$\gamma_{1,1} = 7.75$

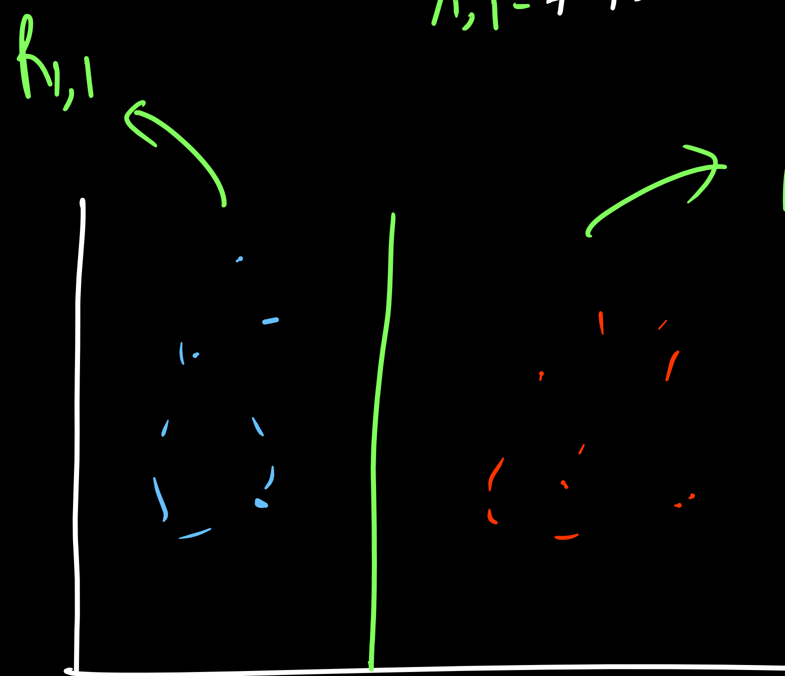
$\gamma_{2,1} = -7.75$

$\gamma_{1,2} = 15.48$

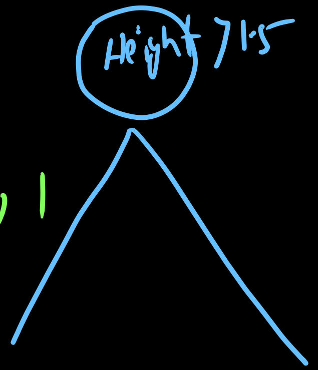
$$\gamma_{2,2} = \frac{-9.97 - 3.97 - 1.52}{3} = -5.15$$

$R_{1,2}$

$R_{2,2}$



$R_{2,1}$



$$\frac{15.48 - 9.97}{2} \qquad \frac{-3.97 - 1.52}{2}$$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ $for\ j = 1, \ldots, J_m$

$$\gamma_{jm} = \operatorname{argmin} \sum_{x_i \in R_{jm}} \left(y^{(i)} - \left(F_{m-1}(x^{(i)}) + \gamma \right) \right)^2$$

$$\rightarrow \quad \frac{\partial}{\partial \gamma} \sum_{x_i \in R_{jm}} \left(y^{(i)} - \left(F_{m-1}(x^{(i)} + \gamma) \right) \right)^2$$

$$\rightarrow \quad \sum_{x_i \in R_{jm}} \left(y^{(i)} - F_{m-1}(x^{(i)}) - \gamma \right) = 0$$

$$\rightarrow \quad \sum_{x_i \in R_{jm}} \left(y^{(i)} - F_{m-1}(x^{(i)}) \right) - \sum_{x_i \in R_{jm}} \gamma = 0$$

$$\rightarrow \quad \sum_{x_i \in R_{jm}} r_{im} = \sum_{x_i \in R_{jm}} \gamma$$

$$= n_j \cdot \gamma$$

✓ $$\boxed{\gamma_{jm} = \frac{1}{n_j} \sum_{x_i \in R_{jm}} r_{im}}$$

Residual;

$x_i \in R_{jm}$

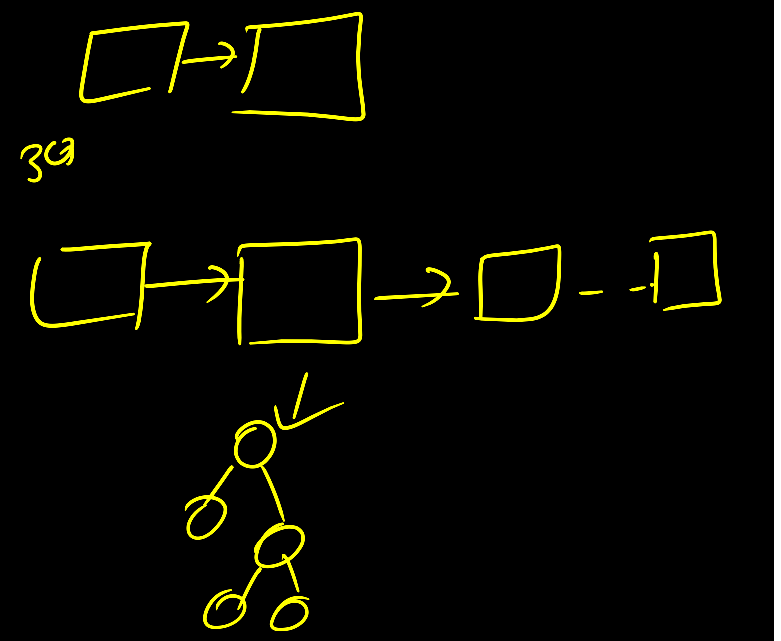
→ Break until 22:27 PM

Hyperparameter

1> M = No. of DT

M↓ → Underfit

M↑ → Overfit

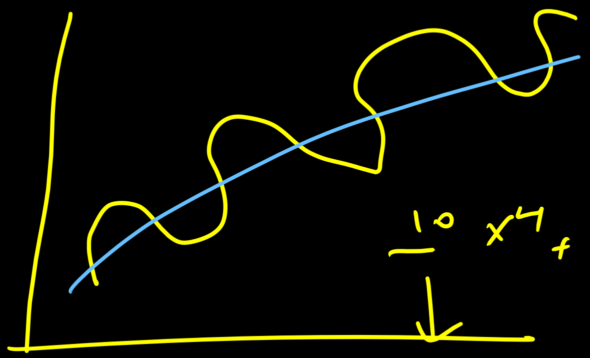


2> Depth ↓ Underfit

↑ Overfit

3> learning rate $\boxed{0.1}$ ✓

✓ $\boxed{1 \rightarrow \text{Overfit}}$



Σw_i^2

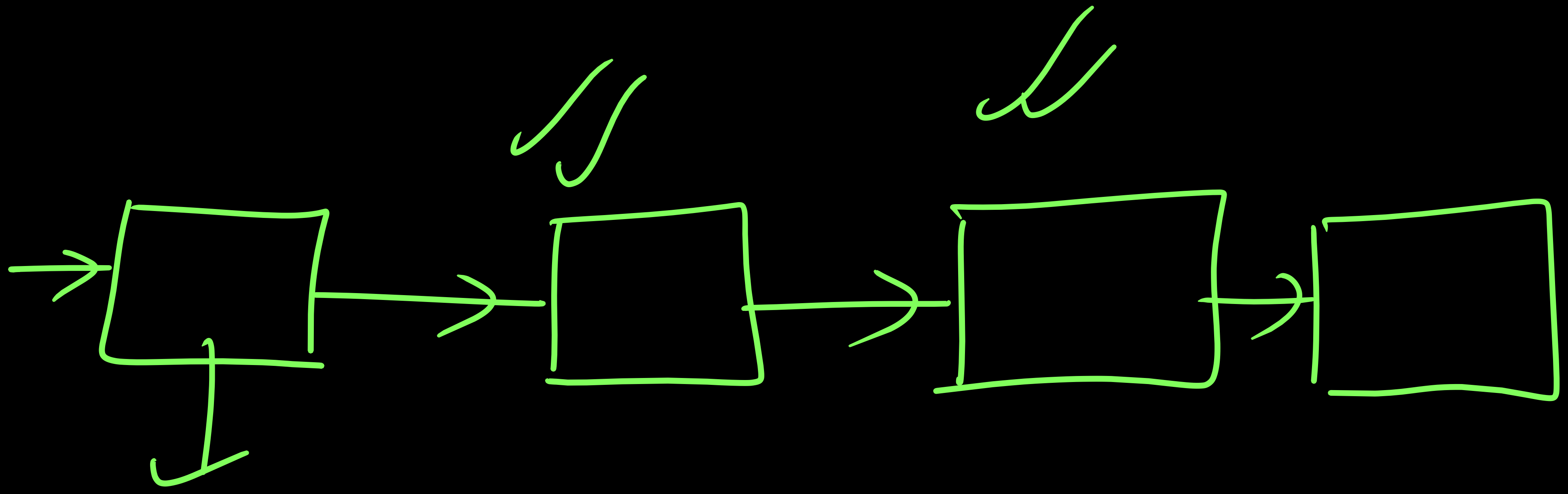
$$\underset{\underset{0.01}{\downarrow}}{\underline{10}}\, x^4 + \underset{\underset{0.1}{\downarrow}}{\underline{15}}\, x^3 + 2x + 1$$

GBDT → Residual + Additive Combining

Stochastic GBDT

→ RS + CS + Residual + Additive Combining

→ Impact of Outliers



Error very
is high for
outlier

→ MSE

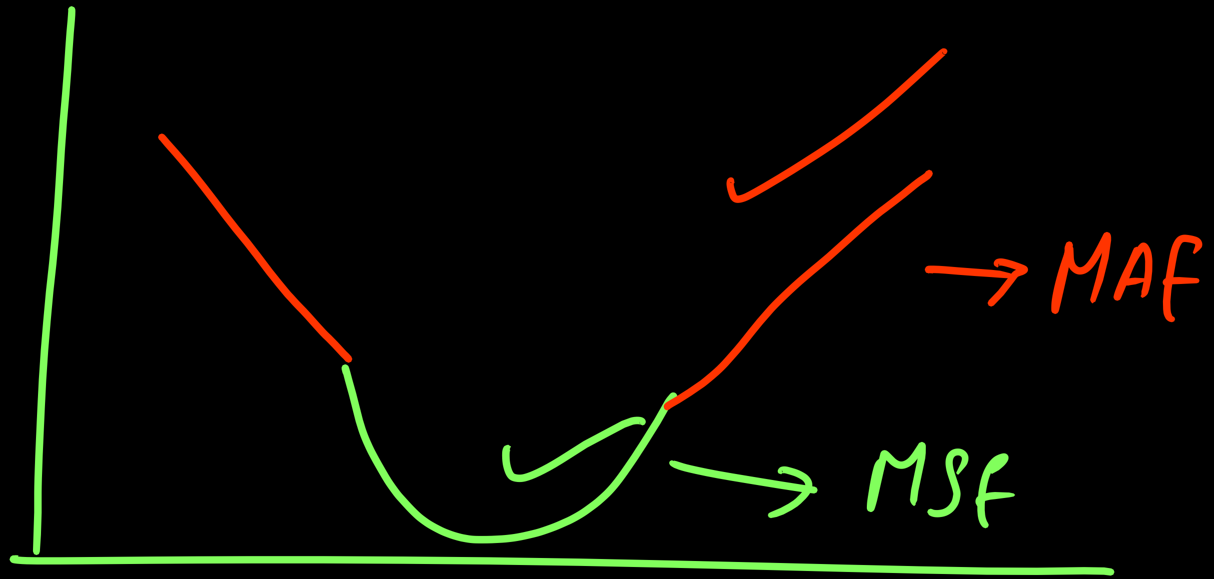
✓ → MAE

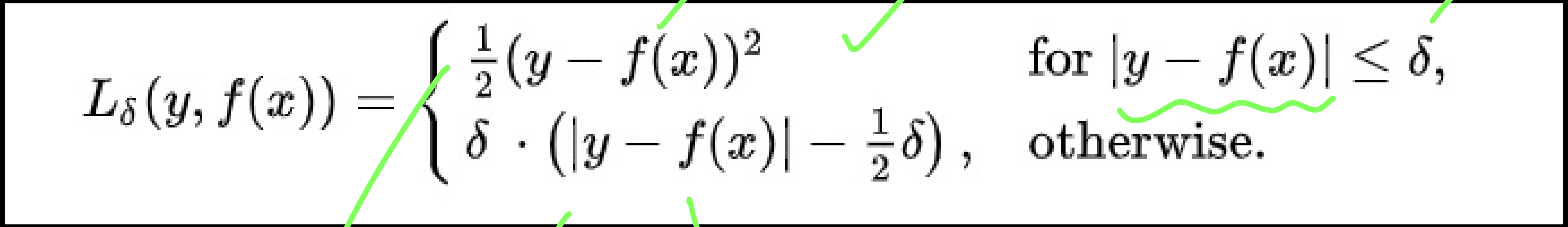
→ Huber loss

↳ MSE + MAE

$$\sum \left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

$$\sum \left|y^{(i)} - \hat{y}^{(i)}\right|$$



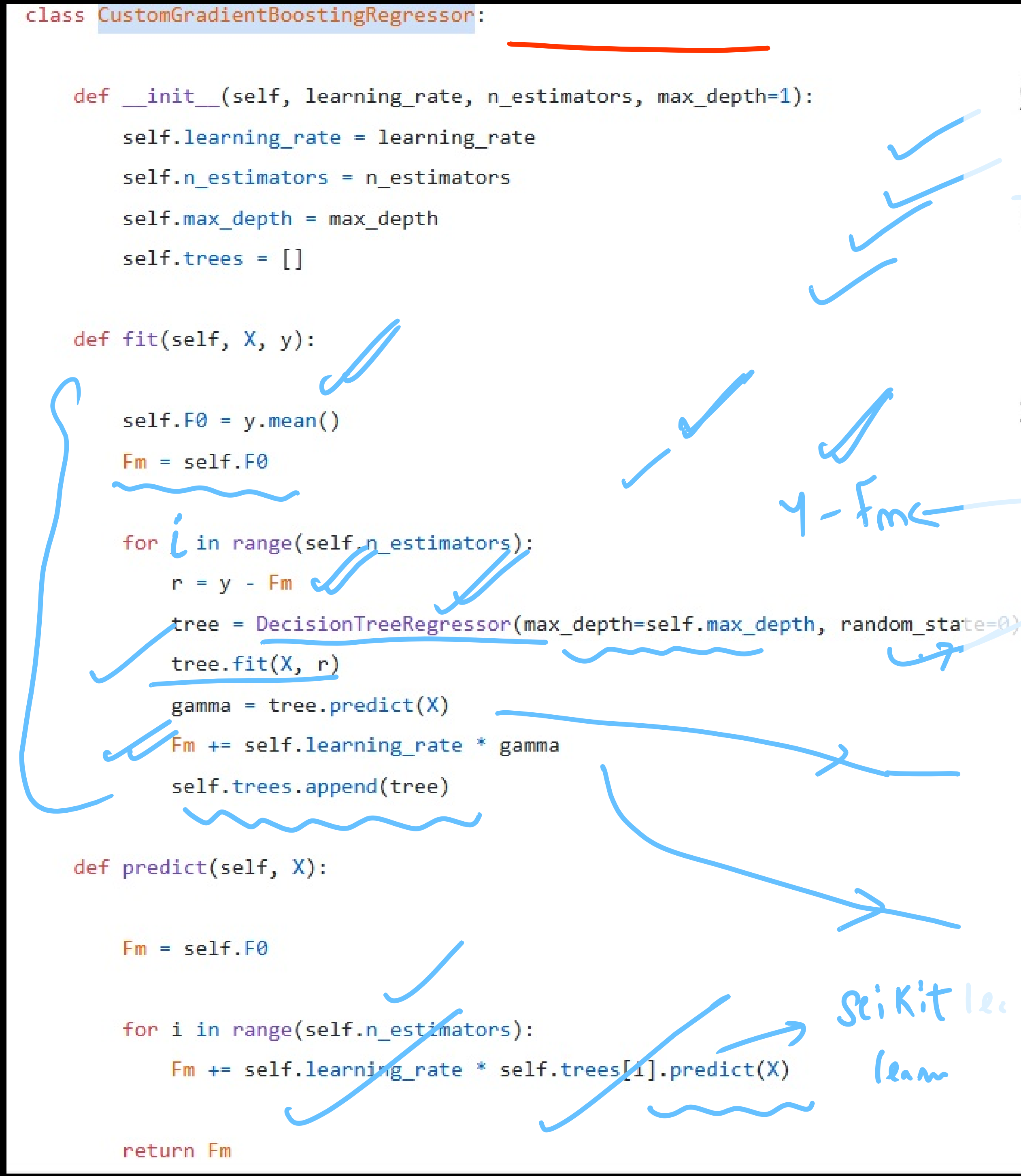


$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \le \delta, \\ \delta \cdot \left(|y - f(x)| - \frac{1}{2}\delta\right), & \text{otherwise.} \end{cases}$$

$\hat{y}$ ✓ 1

MSE ✓ $\hat{y}$

$$1 \cdot \left(|y - \hat{y}| - \frac{1}{2} \cdot 1 \right) \rightarrow \text{MAE}$$

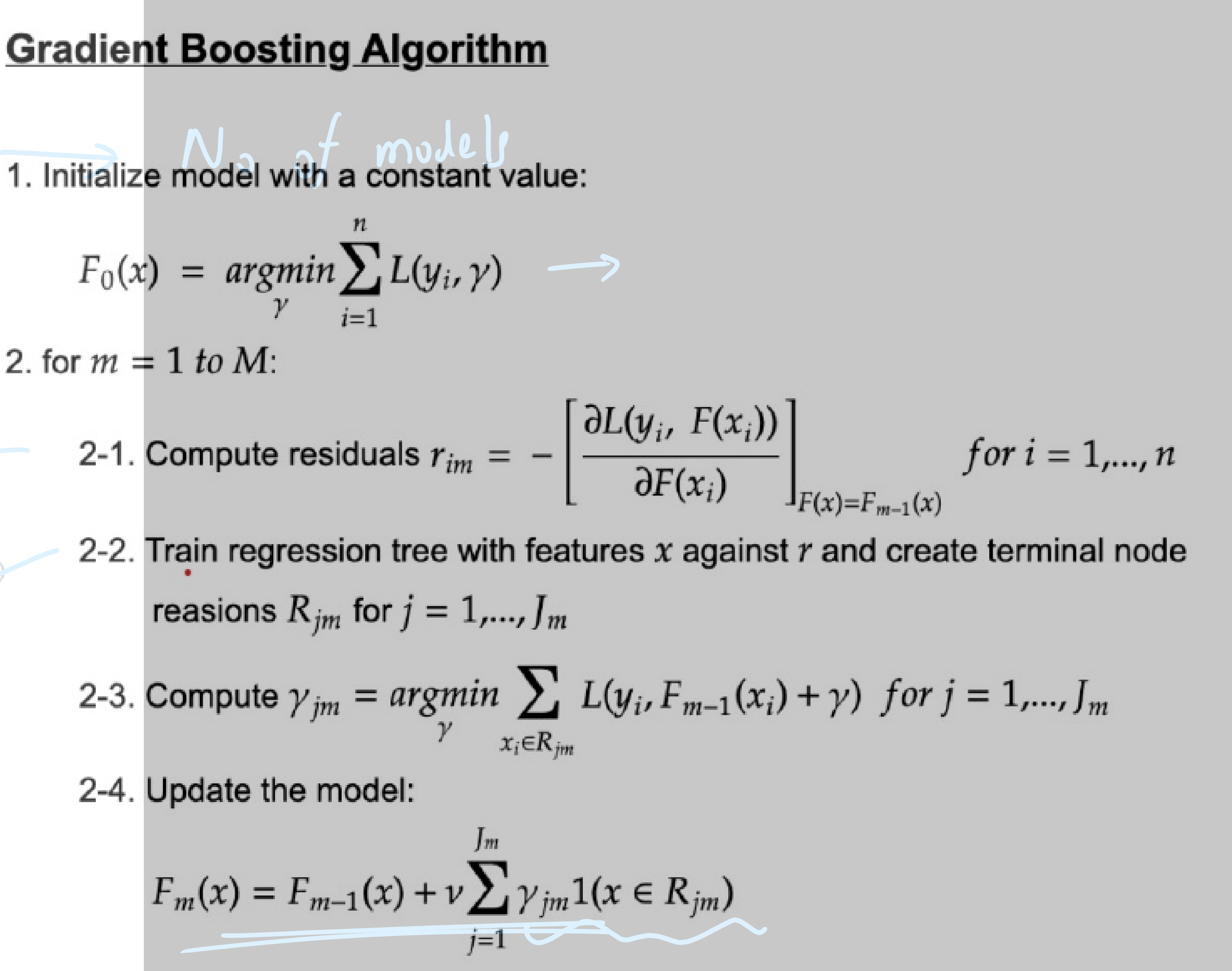


```
class CustomGradientBoostingRegressor:

    def __init__(self, learning_rate, n_estimators, max_depth=1):
        self.learning_rate = learning_rate
        self.n_estimators = n_estimators
        self.max_depth = max_depth
        self.trees = []

    def fit(self, X, y):

        self.F0 = y.mean()
        Fm = self.F0

        for i in range(self.n_estimators):
            r = y - Fm
            tree = DecisionTreeRegressor(max_depth=self.max_depth, random_state=0)
            tree.fit(X, r)
            gamma = tree.predict(X)
            Fm += self.learning_rate * gamma
            self.trees.append(tree)

    def predict(self, X):

        Fm = self.F0

        for i in range(self.n_estimators):
            Fm += self.learning_rate * self.trees[i].predict(X)

        return Fm
```

$y - F_m$

scikit lea

learn



Gradient Boosting Algorithm

No. of models

1. Initialize model with a constant value:

$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$

2. for $m = 1$ to M:

2-1. Compute residuals $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ $for\ i = 1,\dots,n$

2-2. Train regression tree with features x against r and create terminal node reasions R_{jm} for $j = 1,\dots,J_m$

2-3. Compute $\gamma_{jm} = \underset{\gamma}{argmin} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$ $for\ j = 1,\dots,J_m$

2-4. Update the model:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} 1(x \in R_{jm})$$

$F_m = F_0$

$$F_m = F_0 + \gamma \sum_{x \in R_{jm}} y_{jm}$$

$$F_m(x) = f_0(x) + \overset{0.1}{\gamma}\left(\underset{\sim}{\overline{\hat{y}_1}} + \overline{\hat{y}_2} + \overset{\sim}{\hat{y}_3} - - \text{---}\right)$$

learning rate

(overfitting)

Regularization