Last class (15 September)

1) Quizzes + Recap
2) Intro to Boosting
3) Boosting Intuition – How to combine Base Learners?
4) What happens at train & test time
5) GBDT Intuition
6) Sklearn implementation

Today's class

1) Quizzes
2) Recap of prob. & loss fn in Boosted Trees
3) How classification is done in Boosted Trees
4) Bias Variance Trade-off
5) How to regularize GBDT
6) Does Outliers impact GBDT
7) Use Case – EMG Signal Classification

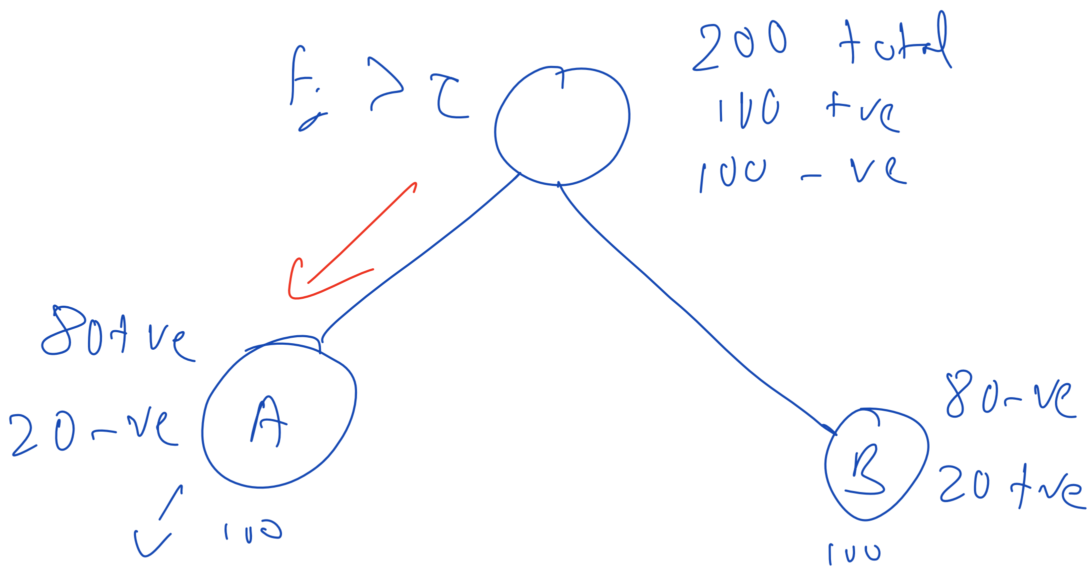
Regression Loss fn $\longrightarrow$ MSE

Classification loss fn $\longrightarrow$ log loss

$$-\sum_{i=1}^{m}\left[y^{(i)}\log\left(\hat{y}^{(i)}\right) + \left(1-y^{(i)}\right)\log\left(1-\hat{y}^{(i)}\right)\right]$$

$y^{(i)} \to$ truth

$\hat{y}^{(i)} \to$ predicted prob



+ve

$p(+ve) = \frac{80}{100} = 0.8$

$p(-ve) = \frac{20}{100} = 0.2$

D : $h_0(x)$ = mean of all y_i

$\hookrightarrow f_0(x)$

$$\underbrace{\boxed{err_1(x^{(i)})}}_{\text{whole data} \;\; h_1(x^{(i)})} = y^{(i)} - f_0(x^{(i)})$$

$$\min_{\gamma} \sum_{i=1}^{m} \Big[\underbrace{y^{(i)}}_{\text{Truth}} - \{ f_0(x^{(i)}) + \gamma h_1(x^{(i)}) \} \Big]^2$$

$$\gamma_1$$

$$f_1(x^{(i)}) = f_0(x^{(i)}) + \gamma_1 h_1(x^{(i)})$$

$$\underbrace{\boxed{err_2(i)}}_{\text{whole data} \;\; h_2(x^{(i)})} = \underbrace{y^{(i)} - f_1(x^{(i)})}_{\text{DT Training}}$$

$$\min_{\gamma} \sum_{i=1}^{m} \Big[y^{(i)} - \{ f_1(x^{(i)}) + \gamma h_2(x^{(i)}) \} \Big]^2 \rightarrow \gamma_2 \qquad \rightarrow L \text{ (loss fn)}$$

$$f_2(x^{(i)}) = f_1(x^{(i)}) + \gamma_2 h_2(x^{(i)})$$

$$\frac{\partial L}{\partial \gamma} = \sum 2\left(y^{(i)} - f_1(x^{(i)}) - \gamma h_2(x^{(i)})\right) \times \left(-h_2(x^{(i)})\right)$$

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

$$2\left[\sum \left(y^{(i)} - f_1(x^{(i)}) - \gamma h_2(x^{(i)})\right) \times h_2(x^{(i)})\right] = 0$$

$$\gamma \sum_{i=1}^{m} h_2^2(x^{(i)}) = \sum \left(y^{(i)} - f_1(x^{(i)})\right) h_2(x^{(i)})$$

$$\Rightarrow \gamma = \frac{\sum_{i=1}^{m} \left(y^{(i)} - f_1(x^{(i)})\right) \cdot h_2(x^{(i)})}{\sum_{i=1}^{m} h_2(x^{(i)}) \times h_2(x^{(i)})}$$

$$f(x^{(q)}) = h_0(x^{(q)}) + \gamma_1 h_1(x^{(q)}) + \gamma_2 h_2(x^{(q)}) + \dots$$

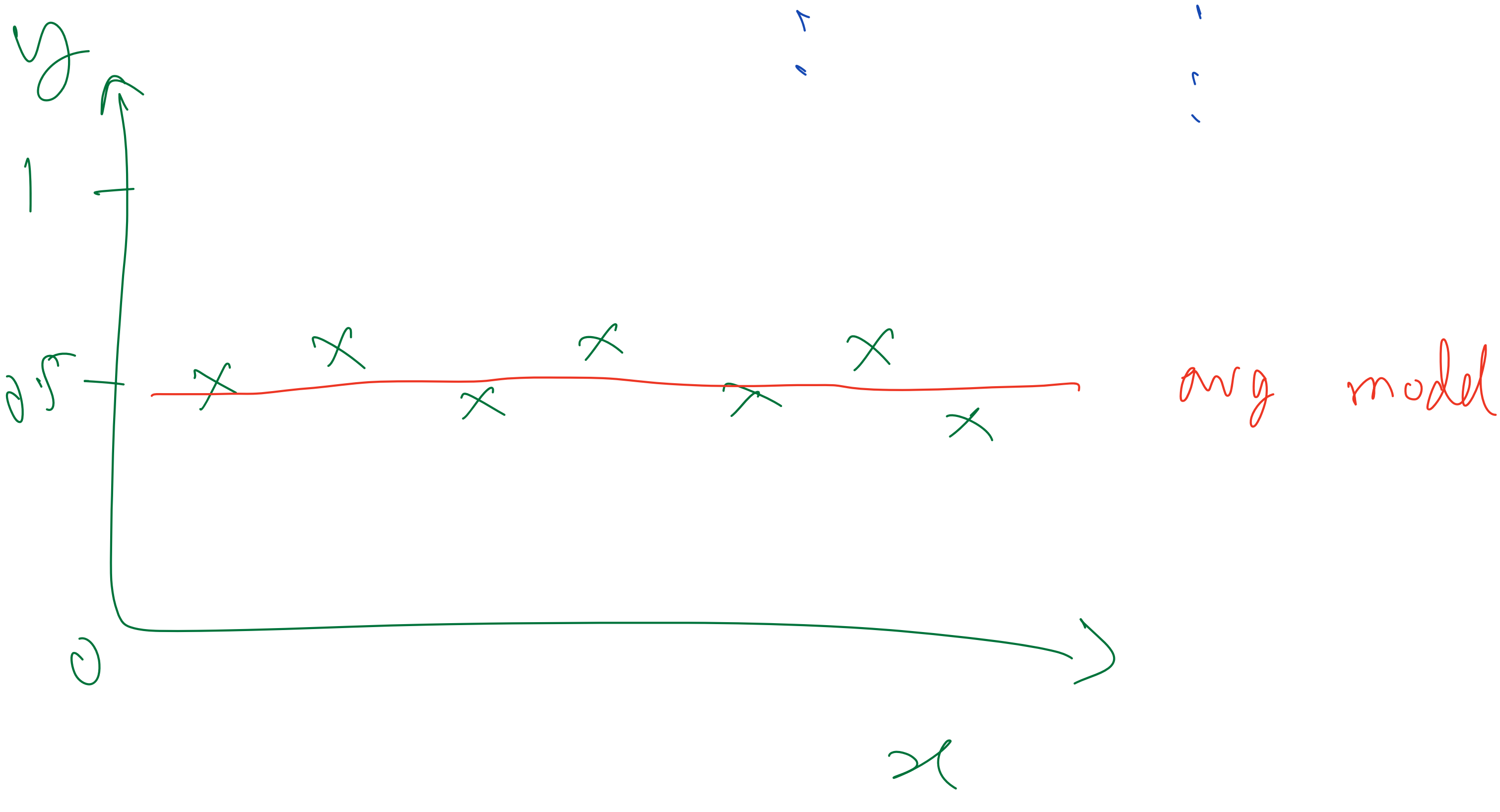
Test time

Training time: $h_0\left(x^{(n)}\right)$

$h_1\left(x^{(n)}\right), \gamma_1$

$h_2\left(x^{(n)}\right), \gamma_2$

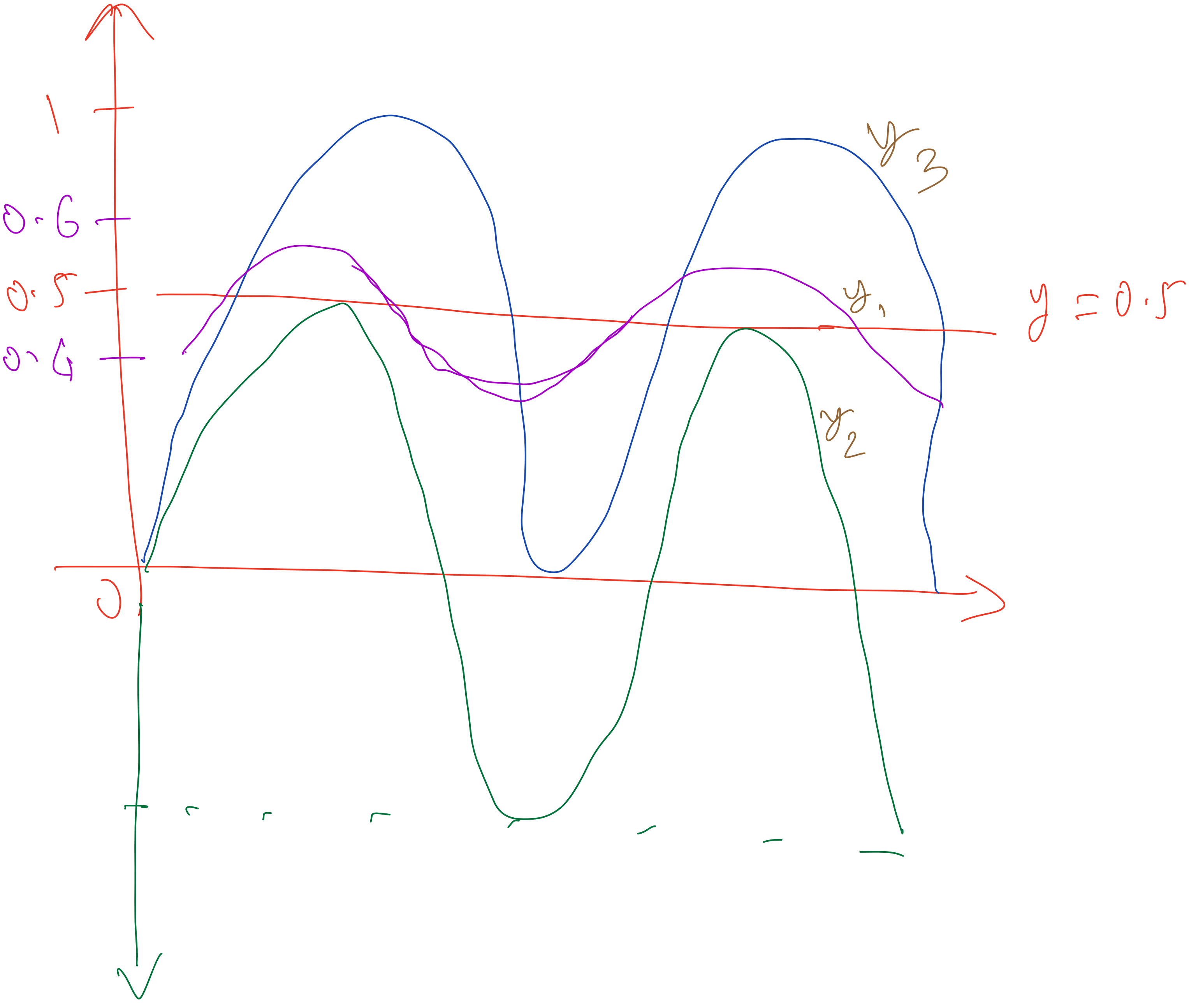
$\vdots \qquad \vdots$



$$\sum \left(y^{(i)} - \hat{y}^{(i)}\right)^2 + \lambda_2 \sum_{i=1}^{d} w_j^2$$

L2 reg coeff.

L2 regulariz

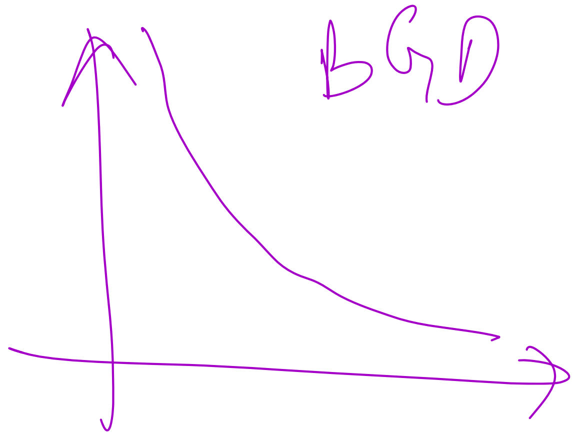


$$y_3 = y_1 + y_2$$

$$y_4 = y_1 + 0.1\, y_2$$

$$W_j = W_j - \eta \times \frac{\partial L}{\partial W_j}$$

learning rate



1000 → 100, 1, 50, 990, . . .

random sequence

10:30 → Resume

shrinkage (ν)

$$F_m(x) = h_0(x) + \boxed{\nu} \sum_{i=1}^{M} \gamma_m h_m(x)$$

Hyper-param only used during training

during validation & test time,

$$F_m(x) = h_0(x) + \sum_{i=1}^{M} \gamma_m h_m(x)$$

$x^{(q)} \rightarrow$ query (test) point

$$\boxed{f_1(x^{(q)})} = h_0(x^{(q)}) + \underbrace{0.1}_{\gamma_1} h_1(x^{(q)})$$

$\boxed{\text{3 classes}}$ → class 0, 1, 2

$$[0.1, 0.5, 0.6] \times [0.7, 0.2, 0.1]$$

?

$$\checkmark\ [0.1, 0.5, 0.4]$$

$$\sum p = 1$$

$$f_1(x^{(i)}) = \begin{bmatrix} 0.1 + 0.1\times 0.7, & 0.5 + 0.1\times 0.2, & 0.4 + 0.1\times 0.1 \end{bmatrix}$$

$$= \begin{array}{ccc} \text{class } 0 & \text{class } 1 & \text{class } 2 \\ \end{array}$$
$$\left[0.17, \boxed{0.52}, 0.41 \right]$$

class 1

$$\gamma_1 = 0.5$$

$$f_1(x^{(i)}) = \begin{bmatrix} 0.1 + 0.5\times 0.7, & 0.5 + 0.5\times 0.2, & 0.4 + 0.5\times 0.1 \end{bmatrix}$$

$$= [0.45, \boxed{0.6}, 0.45]$$

, class 1

normalize to make the sum 1