Quiz 3

Q.1

a) $F_0(x) = \text{mean}(y) = \frac{2+3+4}{3} = 3$

Computing the negative gradient

$$L(y,f) = e^{|y-f|}$$

$$\frac{dL}{df} = e^{|y-f|} \operatorname{sign}(f-y)$$

$$r_i = -\left.\frac{\partial L}{\partial f}\right|_{f=F_0(x_i)}$$

$$= -e^{|y_i - F_0|} \operatorname{sign}(F_0 - y)$$

$$r_i = e^{|y_i - F_0|} \operatorname{sign}(y_i - F_0)$$ 1

for $x=1, y=2$ $\quad r_i = e^{|(2-3)|} (2-3) = -e$

$x=2, y=3$ $\quad r_i = e^{|(3-3)|} (3-3) = 0$

$x=3, y=4$ $\quad r_i = e^{|(4-3)|} (4-3) = +e$

pseudo residuals $r(i) = (e, 0, -e)$ .5

$x = 1.5$ split

$x \leq 1.5$ – (1) $\quad$ residual $= -e$ $\quad$ $h = -e$

$x > 1.5$ – (2,3) $\quad$ residual $= 0, e$ $\quad$ $h = +e/2$ .5

$x = 2.5$ split

$x \leq 2.5$ – (1,2) $\quad$ residual $= -e, 0$ $\quad$ $h = -e/2$

$x > 2.5$ – (3) $\quad$ residual $= e$ $\quad$ $h = e$ .5

Both cuts reduce the squared error by the same amount $(e^2/2)$ so either is acceptable,

Taking the first split

$$h_1(x) = \begin{cases} -e, & x \leq 1.5 \\ e/2, & x > 1.5 \end{cases}$$ .5

in addition as the question asks for exponential-absolute error, check which cut minmizes the exp-abs error. both approaches are fine b)

Update the ensemble (learning rate = 0.1)

$$F_1(x) = F_0(x) + \eta\, h_1(x)$$

$$= 3 + 0.1\, h_1(x)$$ .25

prediction for $x=2$

$2 > 1.5 \quad \therefore \quad h_1(2) = e/2$

$$F_1(2) = 3 + 0.1\,(e/2) = 3 + 0.05e = 3.136$$ .25

CLASSTIME Pg No
Date / /

Q2

$$L(\beta, \beta_0) = e^{(y(\beta^T x + \beta_0))^2} + |\beta_0|$$

Parameters will be updated using gradient descent

$$\beta_0^{new} = \beta_0^{current} - \eta \frac{\partial L}{\partial \beta_0}$$

$$\beta^{new} = \beta^{current} - \eta \frac{\partial L}{\partial \beta}$$

$\eta \to$ learning rate, $x \in \mathbb{R}^n \Rightarrow \beta \in \mathbb{R}^n$; $\beta_0 \in \mathbb{R}$, $y \in \mathbb{R}$

$$\frac{\partial L}{\partial \beta} = e^{(y(\beta^T x + \beta_0))^2} 2y(\beta^T x + \beta_0)\, y x$$ 1

[using chain rule]

$$\frac{\partial L}{\partial \beta_0} = e^{(y(\beta^T x + \beta_0))^2} 2y(\beta^T x + \beta_0) y) + \frac{\partial |\beta_0|}{\partial \beta_0}$$ 1

sign(\beta_0)= $\frac{\partial |\beta_0|}{\partial \beta_0} = \begin{cases} +1 & \beta_0 > 0 \\ -1 & \beta_0 < 0 \\ \text{undefined} & \beta_0 = 0 \end{cases} \Rightarrow$ call this $f(\beta_0)$

.5

d(|x|)/dx at x=0, is in set [-1,1]

final weight update conditions

.5

$$\beta_0^{new} \leftarrow \beta_0^{current} - \eta\left(e^{(y(\beta^T x + \beta_0))^2} 2y(\beta^T x + \beta_0) y + f(\beta_0)\right)$$

$$\beta^{new} \leftarrow \beta^{current} - \eta\left(e^{(y(\beta^T x + \beta_0))^2} 2y(\beta^T x + \beta_0) y x\right)$$

Quiz 3

Ans. 3. $X = [x_1, x_2]^T$

$\beta = [\beta_1, \beta_3, \beta_4, \beta_5]^T$

$h_1 = x_1 \beta_1$
$h_2 = x_2 \beta_3$ } → Hidden layers

Output before activation, $z = h_1\beta_4 + h_2\beta_5 = \beta_1\beta_4 x_1 + \beta_3\beta_5 x_2$ .5

Activation at output node: $\hat{y} = z^2$

∴ loss function, $L(\beta) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}) + \|\beta\|_2^2$

Now, for β_1 using gradient Descent,

$z = \beta_1\beta_4 x_1 + \beta_3\beta_5 x_2 \Rightarrow \hat{y} = z^2$

Using chain rule,

$$\frac{\delta L}{\delta \beta_1} = \frac{\delta L}{\delta \hat{y}} \cdot \frac{d\hat{y}}{dz} \cdot \frac{dz}{d\beta_1} + \frac{\delta}{d\beta_1}\|\beta\|^2$$ .5

$$\therefore \frac{\delta L}{\delta \hat{y}} = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}$$ .5

$$\therefore \frac{\delta \hat{y}}{dz} = 2z$$ .25

$$\therefore \frac{dz}{d\beta_1} = \beta_4 x_1$$ .25

$\therefore \frac{\partial}{\partial \beta_1} \|\beta\|^2 = 2\beta_1$ .5

$\therefore$ Putting all these together,

$$\Rightarrow \frac{\partial L}{\partial \beta_1} = \left(-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \cdot 2z \cdot \beta_4 x_1 + 2\beta_1$$ .5

Now, using Gradient Descent,

$$\beta_1 \leftarrow \beta_1 - \eta \cdot \frac{\partial L}{\partial \beta_1}$$

Putting $\frac{\partial L}{\partial \beta_1}$ from above,

$$\beta_1 \leftarrow \beta_1 - \eta\left[\left(-\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \cdot 2z \cdot \beta_4 x_1 + 2\beta_1\right]$$