Amit Sutradhar

ID: 22201054

Section: 18

Date: 15.05.2025

Assignment

CSE422

Problem - 01

a) Total instances = 14

Yes = 9

No = 5

Entropy, $H(D) = -\left(\frac{9}{14}\log_2\frac{9}{14} + \frac{5}{14}\log_2\frac{5}{14}\right)$

$= 0.9402$

b)

i) Sunny (5 instances):

Yes = 2

No = 3

$H(D_{sunny}) = -\left(\frac{2}{5}\log_2\frac{2}{5} + \frac{3}{5}\log_2\frac{3}{5}\right)$

≈ 0.971

ii) Overcast (4 instances):

Yes = 4

No = 0

$H = 0$

iii) Rainy (5 instances):

Yes = 3

No = 2

$$H(D_{rainy}) = -\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right)$$

$$\approx 0.971$$

$$\therefore H(D|outlook) = \frac{5}{14} \times 0.971 + \frac{4}{14} \times 0 + \frac{5}{4} \times 0.971$$

$$\approx 0.694$$

Information gain for outlook,

$$IG(D, outlook) = H(D) - H(D|outlook)$$

$$= 0.940 - 0.694$$

$$= 0.246$$

c) Resulting entropy for each branch after splitting on Outlook:

Sunny ≈ 0.971

Overcast $= 0$

Rainy ≈ 0.971

Problem - 02

a) Given, $\hat{Y} = w_0 + w_1 x$

Data points = $(1,2), (2,3), (3,5)$

Initial weights, $w_0 = 0$, $w_1 = 0$

$\alpha = 0.1$

Mean squared error (MSE) Loss function,

$$L(w_0, w_1) = \frac{1}{m} \sum_{i=1}^{m} (Y_i - \hat{Y}_i)^2$$

$$= \frac{1}{3} \sum_{i=1}^{3} (Y_i - (w_0 + w_1 x))^2$$

b) The gradients of MSE are,

$$\frac{\partial L}{\partial w_0} = -\frac{2}{3} \sum_{i=1}^{m=3} (Y - (w_0 + w_1 x))$$

$$\frac{\partial L}{\partial w_1} = -\frac{2}{3} \sum_{i=1}^{3} (Y - (w_0 + w_1 x)) x_1$$

Now,

For,

$(1,2)$: $2-(0+0.1)=2$

$(2,3)$: $3-(0+0.2)=3$

$(3,5)$: $5-(0+0.3)=5$

Gradients,

$$\frac{\partial L}{\partial w_0} = -\frac{2}{3}(2+3+5) \approx -6.6667$$

$$\frac{\partial L}{\partial w_1} = -\frac{2}{3}((2\times1)+(3\times2)+(5\times3))$$

$$= -\frac{2}{3}(2+6+15)$$

$$\approx -15.33$$

c) Perform one gradient descent step,

$$w_0^{new} = w_0 - \alpha \frac{\partial}{\partial w_0} L$$

$$= 0 - 0.1 \times (-6.6667)$$

$$= 0.6667$$

$$w_1^{new} = w_{01} - \alpha \frac{\partial}{\partial w_1} L$$

$$= 0 - 0.1 \times (-15.33)$$

$$= 1.533$$

(Ans)

Problem - 03

a) Given, $\hat{y} = \sigma(w_0 + w_1 x_1 + w_2 x_2)$

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad x_1 = 1, \; x_2 = 2, \; y = 1$$

$$w_0 = 0, \; w_1 = 0.5, \; w_2 = -0.5$$

$$z = w_0 + w_1 x_1 + w_2 x_2 = 0 + 0.5 \times 1 + (-0.5) \times 2$$

$$= 0.5 - 1.0 = -0.5$$

$$\hat{y} = \sigma(z) = \frac{1}{1+e^{-(-0.5)}} \approx 0.3775$$

b) binary cross-entropy loss,

$$L = -[y \ln(\hat{y}) + (1-y) \ln(1-\hat{y})]$$

$$\therefore L = -[1 \cdot \log(0.3775) + 0 \cdot \log(1 - 0.3775)]$$

$$= -\ln(0.3775)$$

$$= 0.9741$$

c) compute the gradient of the loss,

$$\frac{\partial L}{\partial w_i} = (\hat{y} - y)x_i$$

(i) for w_0 : $\frac{\partial L}{\partial w_0} = \hat{y} - y = 0.3775 - 1$

$= -0.6225$

(ii) for w_1: $\frac{\partial L}{\partial w_1} = (\hat{y} - y)x_1 = -0.6225 \times 1$

$= -0.6225$

(iii) for w_2: $\frac{\partial L}{\partial w_2} = (\hat{y} - y)x_2 = -0.6225 \times 2$

$= -1.245$

(Ans)

d) weight update rule, $w_i^{new} = w_i - \alpha \frac{\partial L}{\partial w_i}$

update w_0,

$w_0^{new} = 0 - 0.1 \times (-0.6225)$

$= 0.06225$

update ω_1;

$$\omega_1^{new} = 0.5 - 0.1 \times (-0.6225)$$

$$= 0.5 + 0.06225$$

$$= 0.56225$$

update ω_2:

$$\omega_2^{new} = -0.5 - 0.1 \times (-1.295)$$

$$= -0.5 + 0.1295$$

$$= -0.3755$$

updated weights are,

$$\omega_0 \approx 0.06225$$

$$\omega_1 \approx 0.56225$$

$$\omega_2 \approx -0.3755$$

Problem - 04

a) The derivative of binary cross-entropy loss,

$$\frac{\partial L}{da} = -\left(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}}\right)$$

$$= -\frac{1}{0.6} + \frac{0}{0.4}$$

$$\approx -1.6667$$

b) Using chain rule computing the gradient of the loss with respect to ω_1 and ω_2,

we know,

$$\frac{dL}{d\omega_1} = \frac{dL}{d\hat{y}} \cdot \frac{d\hat{y}}{da} \cdot \frac{da}{dz} \cdot \frac{dz}{d\omega_1}$$

(i) $\frac{d\hat{y}}{da} = \hat{y}(1-\hat{y}) \cdot \omega_3 = 0.6(1-0.6) \times 0.4$

$$= 0.096$$

(ii) $a = \sigma(z) = \sigma(0.3) \Rightarrow \frac{da}{dz} = a(1-a)$

$$\sigma(0.3) = \frac{1}{1+e^{-0.3}} \approx 0.5744$$

$$\frac{da}{dz} = 0.5744(1-0.5744)$$

$$\approx 0.2445$$

$$\frac{dz}{d\omega_1} = x_1 = 1 \quad , \quad \frac{dz}{d\omega_2} = x_2 = 2$$

Now,

$$\frac{dL}{d\omega_1} = (-1.6667) \times (0.096) \times (0.2445) \times 1$$

$$\approx -0.0391$$

$$\frac{dL}{d\omega_2} = (-1.6667) \times (0.096) \times (0.2445) \times 2$$

$$\approx -0.0782$$

(Ans)

c)

$\alpha = 0.1$

weight update rule,

$$w_i^{new} = w_i - \alpha \frac{\partial L}{\partial w_i}$$

(i) w_1,

$$w_1^{new} = 0.1 - 0.1 \times (-0.0391)$$

$$\approx 0.1039$$

(ii) w_2,

$$w_2^{new} = 0.2 - 0.1 \times (-0.0782)$$

$$\approx 0.2078$$

updated weights,

$w_1 \approx 0.1039$

$w_2 \approx 0.2078$

(Ans)