孙涛 12232426

1. (1) Overall, maximizing the likelihood with respect to the parameters θ is the same as minimizing the cross-entropy.

$$\prod_{n=1}^{N} P(y^{(n)}|x^{(n)}) = \prod_{n=1}^{N} (1-h_\theta(x^{(n)}))^{y^{(n)}} (h_\theta(x^{(n)}))^{(1-y^{(n)})}$$

$$-\frac{1}{N}\log\left(\prod_{n=1}^{N} P(y^{(n)}|x^{(n)})\right) = -\frac{1}{N}\sum_{n=1}^{n} y(n)\log(1-h_\theta(x^{(n)})) + (1-y^{(n)})\log(h_\theta(x^{(n)}))$$

(let the base of log is e)

$$= \frac{1}{N}\sum_{n=1}^{n}[y^{(n)}\ln(1-h_\theta(x^{(n)})) - (1-y^{(n)})\ln(h_\theta(x^{(n)}))]$$

$$= L_\theta$$

(2)
$$\frac{\partial L_\theta}{\partial \theta} = -\frac{1}{N}\sum_{n=1}^{N} \frac{(-y^{(n)}) \cdot \frac{\partial h_\theta(x^{n})}{\partial \theta}}{1-h_\theta(x^{n})} + \frac{(1-y^{(n)}) \cdot \frac{\partial h_\theta(x^{(n)})}{\partial \theta}}{h_\theta(x^{(n)})}$$

$$= -\frac{1}{N}\sum_{n=1}^{N} \frac{\partial h_\theta(x)}{\partial \theta}\left(\frac{-y^{(n)}}{\frac{e^{-f_\theta(x)}}{1+e^{-f_\theta(x)}}} + \frac{1-y^{(n)}}{\frac{1}{1+e^{-f_\theta(x^{n})}}}\right)$$

$$= -\frac{1}{N}\sum_{n=1}^{N} \left(\frac{1}{1+e^{-f_\theta(x^{n})}}\right)^2 e^{-f_\theta(x^{(n)})}\frac{\partial f_\theta(x^{(n)})}{\partial \theta}\cdot\left(\frac{-y^{(n)}\cdot(1+e^{-f_\theta(x^{(n)})})}{e^{-f_\theta(x^{(n)})}} + (1-y^{(n)})(1+e^{-f_\theta(x^{(n)})})\right)$$

$$= -\frac{1}{N}\sum_{n=1}^{N}\left[\left(y^{(n)} - \frac{e^{-f_\theta(x^{(n)})}}{1+e^{-f_\theta(x^{(n)})}}\right)\cdot\frac{\partial f_\theta(x^{(n)})}{\partial \theta}\right]$$

2. (1) $h_1 = \text{sigmoid}(0.2 + 0.1 \times 0.2 + 0.15 \times 0.3) \approx 0.5659$

$h_2 = \text{sigmoid}(0.2 + 0.1 \times 0.15 + 0.15 \times 0.25) \approx 0.5628$

$o_1 \overset{\text{sigmoid}}{=} (0.4 + h_1 \times 0.4 + h_2 \times 0.6) \approx 0.7239$

$o_2 = \text{sigmoid}(0.4 + 0.35 \times h_1 + 0.1 \times h_2) \approx 0.6580$

That is $out_{o_1} = 0.7239$ $out_{o_2} = 0.6580$

(2) $E_{total} = (o_1 - 0.99)^2 + (o_2 - 0.01)^2$

$= (0.7239 - 0.99)^2 + (0.6580 - 0.01)^2$

$= 0.4907$ The $MSE = 0.4907 \times \frac{1}{2} = 0.2454$

(3) $\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial o_1} \cdot \frac{\partial o_1}{\partial w_5} = \frac{\partial \frac{1}{2}(o_1 - y_1)^2}{\partial o_1} \cdot o_1 \cdot (1 - o_1) \cdot h_1$

$= (o_1 - y_1) o_1 \cdot (1 - o_1) \cdot h_1$

when $\eta = 0.1$

$w_5^{new} = w_5 - \eta \cdot \frac{\partial E_{total}}{\partial w_5}$

$= 0.4 - 0.1 \times (0.7239 - 0.99) \cdot 0.7239 \cdot (1 - 0.7239) \cdot 0.5659$

$= 0.403$

S

3. Information gain $IG(X) = H(Y) - H(Y|X)$ $H(Y) = -(\frac{1}{3}\cdot\log_2\frac{1}{3})\cdot 3 = \cancel{0.4771}$ 1.585

Since the initial $H(Y)$ is the same, we could select the tree root based on $H(Y|X)$. $H(X) = -\sum_k p(y_k)\log_2 p(y_k)$

known that, $H(Y|X) = \sum_j p(X=x_j)\cdot H(Y|X=x_j)$

① height ($x \le 175$, $175 < x < 183$, $x \geq 183$)

$$H(Y|\text{height}) = \frac{3}{9}\times 0 + \frac{2}{9}\times 0 + \frac{4}{9}\left(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4}\right)$$

$$= \frac{1}{9}\log_2 4 + \frac{3}{9}\log_2\frac{4}{3}$$

$$= 0.3603$$

② weight ($x < 70$, $70 \le x \le 80$, $x > 80$)

$$H(Y|\text{weight}) = \frac{3}{9}\times 0 + \frac{3}{9}\times\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right) + \frac{3}{9}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right)$$

$$= \frac{2}{9}\log_2 3 + \frac{4}{9}\log_2\frac{3}{2}$$

$$= 0.6122$$

③ eye-color (hazel, brown, blue)

$$H(Y|\text{eye-color}) = \frac{1}{3}\left(-3\cdot\frac{1}{3}\log_2\frac{1}{3}\right) + \frac{1}{3}\left(-3\cdot\frac{1}{3}\log_2\frac{1}{3}\right) + \frac{1}{3}\left(-3\cdot\frac{1}{3}\log_2\frac{1}{3}\right)$$

$$= \log_2 3 = 1.585$$

④ hair-color (black, blond, brown)

$$H(Y|\text{hair-color}) = \frac{1}{3}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}\right)\times 3$$

$$= \frac{1}{3}\log_2 3 + \frac{2}{3}\log_2\frac{3}{2}$$

$$= 0.9183$$

Among all the results above, we can know that choose height as the tree's root. In this way, we will obtain maximum information gain