$$\mathcal{L}_i = \frac{1}{2}(y_i - t_i)^2 \qquad J = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i$$

$$J = \frac{1}{2N}\sum_{i=1}^{N}(y_i - t_i)^2 = \frac{1}{2N}\sum (w^T x^i + b - t^i)^2$$

$\nabla J_w = \left(\frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2}, \ldots, \frac{\partial J}{\partial w_D}\right)$. D dimensionalities for each $x_i$

$$-P\log_2 P - (1-P)\log_2(1-P)$$

$$H(Y) = -\sum_{y \in Y} P(y)\log_2 P(y)$$

$$-\frac{8}{9}\log_2\frac{8}{9} - \frac{1}{9}\log_2\frac{1}{9} = 0.5$$

$$-\frac{4}{9}\log_2\frac{4}{9} - \frac{5}{9}\log_2\frac{5}{9} = 0.99$$

| X \ Y | cloudy | ¬cloudy |
|---|---|---|
| rain | 24/100 | 1/100 |
| ¬rain | 23/100 | 50/100 |

$$H(X,Y) = -\sum_{x \in X}\sum_{y \in Y} P(x,y)\log_2 P(x,y)$$

$$= -(P(0,0)\log_2 P(0,0) + P(0,1)\log_2 P(0,1) + P(1,0)\log_2 P(1,0) + P(1,1)\log_2 P(1,1))$$

$$= -\left(\frac{50}{100}\log\frac{50}{100} + \frac{23}{100}\log_2\frac{23}{100} + \frac{1}{100}\log_2\frac{1}{100} + \frac{24}{100}\log_2\frac{24}{100}\right) = 1.56$$

$$H(Y|X=x) = -\sum_{y \in Y} P(y|x) \log_2 P(y|x)$$

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

$$= -\left( \frac{P(\text{cloudy} \ddagger \text{rain})}{P(\text{rain})} \log_2 \frac{P(\text{cloudy} \ddagger \text{rain})}{P(\text{rain})} + \frac{P(\neg\text{cloudy} \ddagger \text{rain})}{P(\text{rain})} \log_2 \frac{P(\neg\text{cloudy} \ddagger \text{rain})}{P(\text{rain})} \right)$$

$$= -\left( \frac{24}{100} \times \frac{100}{25} \log \frac{24}{25} + \frac{1}{100} \times \frac{100}{25} \log_2 \frac{1}{25} \right) = 0.24 \text{ bits.}$$

$$H(Y|X) = \sum_{x \in X} P(x) H(Y|X=x) = -\sum_{x \in X} \sum_{y \in Y} \underbrace{P(x) P(y|x)}_{P(x,y)} \log P(y|x)$$

$$H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} P(x,y) \log_2 P(y|x)$$

$$= -\sum_{x} \sum_{y} P(x,y) \log_2 P(y|x)$$

$$P(y|x) = \frac{P(x,y)}{P(x)} \Rightarrow P(y|x) \cdot P(x) = P(x,y)$$

$$H(Y|X) = \sum_{x \in X} P(x) H(Y|X=x)$$

$$= \frac{1}{4} H(Y = \text{cloudiness} \mid X = \text{rain}) + \frac{3}{4} H(Y = \text{cloudiness} \mid X = \neg\text{rain})$$

$$= \frac{1}{4} \cdot 0.24 + \frac{3}{4} \cdot ? = 0.75 \text{ bits}$$

$$-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0.86$$

$$IG = 0.86 - \left( 0.81 \cdot \frac{4}{7} + 0.92 \cdot \frac{3}{7} \right) = 0.006$$

$$H(Y|\text{left}) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 0.81 \quad / \frac{4}{7}$$

$$H(Y|\text{right}) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.92 \quad / \frac{3}{7}$$