## 1-1

$H(X) = \sum_{x\in X} -p(x)\log_2(p(x))$.

1) 공정한 주사위.

$6\left(\frac{1}{6}\log_2 6\right) = \boxed{\log_2 6}$

2) 모든 눈이 6인 주사위

$\boxed{1\log_2 1 = 0}$

3) 짝수 눈이 나올 확률이 홀수 눈이 나올 확률의 2배인 주사위.

$\underbrace{(a+a+a)}_{3a}+\underbrace{(2a+2a+2a)}_{6a}=1$. $\quad a=\frac{1}{9}$ $\quad 3\left(\frac{1}{9}\log_2 9\right)+3\left(\frac{2}{9}\log_2\frac{9}{2}\right)$

$9a=1$ $\quad \frac{1}{3}\log_2 9 + \frac{2}{3}(\log_2 9 - \log_2 2) = \boxed{\log_2 9 - \frac{2}{3}}$

4) 1이 나올 확률이 0.5, 나머지 눈이 나올 확률이 각각 0.1.

$\frac{1}{2}\log_2(2) + 5\left(\frac{1}{10}\log_2 10\right) = \frac{1}{2}+\frac{1}{2}\log_2 10 = \log_2\sqrt{20}$

$\frac{1}{2}(1+\log_2 10)$

$\log_2 20$

답은 1번

## 1-2

$H(X)= \sum_{x\in X} -p(x)\log_2\{p(x)\}$.

1) $0\le H(x)$ 증명

$H(X)= \sum_{x\in X} \underbrace{-p(x)}_{\text{양수}}\,\underbrace{\log_2\{p(x)\}}_{\text{음수}}$.

↳ 양수 ↳ 음수. (∵ $\log_2(0,1)$ 꼴이니).

2) $H(x)\le\log_2|x|$

최대 엔트로피는 균등분포에서 발생한다.

$H(X)= -\sum_{i=1}^{n}\frac{1}{n}\cdot\log_2\frac{1}{n} = -n\cdot\frac{1}{n}\log_2\frac{1}{n} = \log_2 n$ $\quad \boxed{H(X)\le\log_2|X|}$.

# 1-3

$$H(X) = \sum_{x \in X} p(x) \log_2 \left(\frac{1}{p(x)}\right).$$

1)

| | |
|---|---|
| $X=0$ | 0.6 |
| $X=1$ | 0.4 |

$$H(X) = \frac{6}{10}\log_2\frac{10}{6} + \frac{4}{10}\log_2\frac{10}{4}$$

$$= 0.99277 \text{ bits}$$

| $Y=0$ | $Y=1$ |
|---|---|
| 0.55 | 0.45 |

$$H(Y) = \frac{55}{100}\log_2\frac{100}{55} + \frac{45}{100}\log_2\frac{100}{45}$$

$$= 0.99277 \text{ bits.}$$

2) $H(X,Y)$

$$\frac{45}{100}\log_2\frac{100}{45} + \frac{15}{100}\log_2\frac{100}{15} + \frac{1}{10}\log_2 10 + \frac{3}{10}\log_2\frac{10}{3} = 1.78233 \text{ bits.}$$

3) $H(X|Y)$

$$H(X|Y) = H(X,Y) - H(Y) = 1.78223 - 0.99277 = 0.78945 \text{ bits.}$$

4) $I(X;Y)$

$$H(X) - H(X|Y) = 0.97095 - 0.78945 = 0.1815 \text{ bits.}$$

5)

$I(X;Y) = H(X) - H(X|Y)$ 란, Y(흡연 여부)를 알았을 때 X(폐암 발생 여부)에 대한 불확실성이 얼마나 줄어드는지를 나타낸다.

$I(X;Y) = 0$ → Y를 알아도 X에 대한 불확실성이 줄어들지 않는다.

$I(X;Y) = H(X)$ → Y를 알면 X는 무조건 안다.

2-1

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}.$$

$$A - \lambda I = \begin{bmatrix} 2-\lambda & 0 & 0 \\ 1 & 2-\lambda & 1 \\ -1 & 0 & 1-\lambda \end{bmatrix} \qquad \det(A-\lambda I) = (2-\lambda)^2(1-\lambda) \qquad \lambda = 2, 1$$

① $\lambda=2$인 경우

$$(A-2I)\vec{v} = 0 \Rightarrow \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ -1 & 0 & -1 \end{bmatrix}\vec{v} = 0 \qquad \vec{v_1} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \vec{v_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

② $\lambda=1$인 경우

$$(A-I)\vec{v} = 0 \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 0 \end{bmatrix}\vec{v} = 0 \qquad \vec{v_3} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

$$P = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2-3)

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = U\Sigma V^T$$

$U$: $BB^T$의 고유벡터들로 구성된 2X2 직교행렬.

$\Sigma$: 특이값을 대각 원소로 갖는 2X3 행렬.

$V$: $B^TB$의 고유벡터들로 구성된 3X3 직교행렬

$$B^TB = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

특성방정식: $\det(B^TB - \lambda I) = 0$ $\quad \lambda_1 = 2, \lambda_2 = 0, \lambda_3 = 1$

$\sigma_1 = \sqrt{2} \quad \sigma_2 = 1 \quad \sigma_3 = 0.$

$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

# 3-1

$$J(W) = \sum_{i=1}^{m} [-y^{(i)} \log \hat{y}^{(i)} - (1-y^{(i)}) \log (1-\hat{y}^{(i)})].$$

$\hat{y}^{(i)} = \sigma(W^T x^{(i)}) = \frac{1}{1+e^{-W^T x^{(i)}}}$ 가 W에 대해 Convex임을 보이자.

$\hat{y}^{(i)}{}' = \sigma''(W^T x^{(i)})$

$\sigma(z) = \frac{1}{1+e^{-z}} \Rightarrow \hat{y}^{(i)} = \frac{1}{1+e^{-W^T x^{(i)}}}$

$l(W) = -y \log(\hat{y}) - (1-y) \log(1-\hat{y}) \Rightarrow l(W) = l(\sigma(W^T x))$

$\nabla_W l(W) = (\hat{y} - y) x$

$\nabla_W^2 l(W) = \hat{y}(1-\hat{y}) \cdot x x^T \geq 0 \Rightarrow$ Convex 함수.

# 3-2

$f(x) = x^4 - 2x^3 - 3x^2 + x$

$x_0 = 1, \ \gamma = 0.2$

$x_{t+1} = x_t - \gamma \nabla f(x_t)$

$f'(x) = 4x^3 - 6x^2 - 6x + 1$

$x_1 = x_0 - \gamma f'(x_0)$ where $x_0 = 1$.

$f'(1) = 4 - 6 - 6 + 1 = -7 \Rightarrow x_1 = 2.4$.

$f'(2.4) = 4(2.4)^3 - 6(2.4)^2 - 6(2.4) + 1 = 7.336$

$x_2 = 2.4 - 0.2 \cdot 7.336 = \boxed{0.9328}$.

# 3-3

$\min f(x,y) = x^2+y^2 \quad s.t.\ x+y=2.$

Step1. 라그랑지안 함수 설정

$L(x,y,\lambda) = f(x,y) + \lambda \cdot g(x,y) = x^2+y^2+\lambda(x+y-2)$

Step2. 편미분 후 stationary point 조건 설정.

$\frac{\partial L}{\partial x} = 2x+\lambda=0 \longrightarrow \lambda=-2x$

$\frac{\partial L}{\partial y} = 2y+\lambda=0 \longrightarrow \lambda=-2y$

$\rightarrow x=y$

$\frac{\partial L}{\partial \lambda} = x+y-2=0$

$\longrightarrow x=y=1.$

$f(x,y) = x^2+y^2 = 1^2+1^2=2.$

stationary point $(x,y,\lambda) = (1,1,-2)$

최소값 : 2.