1.

(1) $\det(A) = \sum_{j=1}^{3} (-1)^{1+j} \cdot a_{1j} \det A_{1j}$

$= 2 \cdot \det\left(\begin{bmatrix} 2 & 3 \\ 0 & 3 \end{bmatrix}\right) - 1 \cdot \det\left(\begin{bmatrix} 0 & 3 \\ 0 & 3 \end{bmatrix}\right) + 1 \cdot \det\left(\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}\right)$

$= 2 \cdot (6-0) - (0-0) + (0-0) = 12$

⇒ 3X3 행렬의 determinant는 일종의 부피다!

(2) $Av = \lambda v$

$(A - \lambda I)v = 0$

⇒ $\det(A - \lambda I) = 0$ ($\because$ v is not a zero vector)

$|A - \lambda I| = \begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & 2-\lambda & 3 \\ 0 & 0 & 3-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda)(3-\lambda) = 0$

$\Rightarrow \lambda = 1, 2, 3$

$(A-I)v_{\lambda=1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \Rightarrow \begin{cases} v_3 = 0 \\ v_2 + 3v_3 = 0 \\ 2v_3 = 0 \end{cases} \Rightarrow v_{\lambda=1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$(A-2I)v_{\lambda=2} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \Rightarrow \begin{cases} -v_1 + v_3 = 0 \\ 3v_3 = 0 \\ v_3 = 0 \end{cases} \Rightarrow v_{\lambda=2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$(A-3I)v_{\lambda=3} = \begin{bmatrix} -2 & 0 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \Rightarrow \begin{cases} -2v_1 + v_3 = 0 \\ -v_2 + 3v_3 = 0 \\ 0 = 0 \end{cases} \Rightarrow v_{\lambda=3} = \begin{bmatrix} \frac{1}{2} \\ 3 \\ 1 \end{bmatrix}$

⇒ 고유벡터는 선형변환에 의해 방향이 바뀌지 않는 벡터, 고유값은 벡터의 크기가 변한 정도이다!

2.

$Bv = \lambda v$

$(B - \lambda I)v = 0$

$|B - \lambda I| = \begin{vmatrix} -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 3 & 0 & 1-\lambda \end{vmatrix} = \lambda^2(1-\lambda) = 0$

$\Rightarrow \lambda = 0, 1$

$\Rightarrow D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$B v_{\lambda=0} = 0 \Rightarrow \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \Rightarrow 3v_1 + v_3 = 0 \Rightarrow v_{\lambda=0} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}$

$(B-I)v_{\lambda=1} = 0 \Rightarrow \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \Rightarrow \begin{cases} -v_1 = 0 \\ -v_2 = 0 \\ 3v_1 = 0 \end{cases} \Rightarrow v_{\lambda=1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\Rightarrow P = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix} \Rightarrow \det(P) = 1 \Rightarrow P^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix}$

$\Rightarrow B = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix} = PDP^{-1}$

3.

$$P(\text{Spam} \mid \text{'당첨'}) = \frac{P(\text{spam}, \text{'당첨'})}{P(\text{'당첨'})} = \frac{P(\text{'당첨'} \mid \text{spam})\, P(\text{spam})}{P(\text{'당첨'} \mid \text{spam}) \times P(\text{spam}) + P(\text{'당첨'} \mid \text{Not spam}) \times (1 - P(\text{spam}))} = \frac{0.5 \times 0.2}{0.5 \times 0.2 + 0.01 \times 0.8} \approx 0.9259$$

4.

(2) Let X be the hit or out and $X \sim P$, then

$$H(X) = \sum_{x \in \mathcal{X}} P(x) \log_2 \frac{1}{P(x)} = \frac{9}{15} \log_2 \frac{15}{9} + \frac{6}{15} \log_2 \frac{15}{6}$$

$$= \frac{3}{5}(\log_2 5 - \log_2 3) + \frac{2}{5}(\log_2 5 - 1)$$

$$= \log_2 5 - \frac{3}{5}\log_2 3 - \frac{2}{5}$$

(3) Compute the KL divergence between P and Q, i.e. $Q(x) = \begin{cases} \frac{4}{15}, & x \text{ is hit} \\ \frac{11}{15}, & x \text{ is out} \end{cases}$

$$KL(P \| Q) = \sum_{x \in \mathcal{X}} P(x) \log_2 \frac{P(x)}{Q(x)} = \frac{9}{15} \log_2 \frac{9}{4} + \frac{6}{15} \log_2 \frac{6}{11}$$

$$= \frac{6}{5}\log_2 3 - \frac{6}{5} + \frac{2}{5} + \frac{2}{5}\log_2 3 - \frac{2}{5}\log_2 11$$

$$= \frac{8}{5}\log_2 3 - \frac{2}{5}\log_2 11 - \frac{4}{5}$$

5.

Claim $J(w) = \sum_{i=1}^{m} -y^{(i)} \log \frac{1}{1+e^{-w^T x^{(i)}}} - (1-y^{(i)}) \log \frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}}$ is convex.

non-negative

Note that $-\log \frac{e^{-w^T x^{(i)}}}{1+e^{-w^T x^{(i)}}} = w^T x^{(i)} - \log \frac{1}{1+e^{-w^T x^{(i)}}}$

It is enough to show that $\underbrace{-\log \frac{1}{1+e^{-w^T x^{(i)}}}}_{:= f(w)}$ is convex in w to prove J(w) is convex.

Since $\nabla^2 f(w) \succeq 0$,

J(w) is convex in w.

Then, we can say that $\nabla J(w) = 0$ implies $w = w^*$ where $w^* = \min_w J(w)$

+ GD Algorithm

① Randomly initialize $w^{(0)}$

② For $t = 0, 1, \cdots, T-1$

$w^{(t+1)} \leftarrow w^{(t)} - \eta \nabla J(w^{(t)})$

③ Return $w^{(T)}$