Q1. $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{pmatrix}$

(1) $\det(A) = 1 \times 6 = 6$

(2) $\det(\lambda I - A) = \det \begin{pmatrix} \lambda-1 & 0 & -1 \\ 0 & \lambda-2 & -3 \\ 0 & 0 & \lambda-3 \end{pmatrix} = (\lambda-1)(\lambda-2)(\lambda-3) \overset{\text{let}}{=} 0 \Rightarrow$ eigenvalue $\lambda = 1, 2, 3$

(3) i) $\lambda = 1$ $\begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & -3 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$x_3 = 0$

$-x_2 - 3x_3 = 0 \rightarrow x_2 = 0.$ $x_1 = 1$

eigenvector corresponding to eigenvalue $\lambda = 1$: $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

ii) $\lambda = 2$ $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & -3 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$x_1 - x_3 = 0$ $x_3 = 0.$ $x_1 = 0$ $x_2 = 1.$

eigenvector corresponding to eigenvalue $\lambda = 2$: $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

iii) $\lambda = 3$ $\begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$2x_1 - x_3 = 0$

$x_2 - 3x_3 = 0$

let $x_2 = 6 \rightsquigarrow x_3 = 2,\ x_1 = 1$

eigenvector corresponding eigenvalue $\lambda = 3$: $\begin{pmatrix} 1 \\ 6 \\ 2 \end{pmatrix}$

Q2. $B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{pmatrix}$

$\det(I\lambda - B) = \det \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -3 & 0 & \lambda-1 \end{pmatrix} = \lambda^2(\lambda-1) \overset{\text{let}}{=} 0.$ $\lambda = 0, 1.$

i) $\lambda = 0.$ $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -3 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $-3x_1 - x_3 = 0.$ $-3x_1 = x_3$ eigenvector : $\begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$ $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

ii) $\lambda = 1.$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ $x_1 = 0.$ $x_2 = 0$ $x_3 = 1.$ eigenvector: $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$

Diagonalization : $\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}}_{P^{-1}} \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{pmatrix}}_{B} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}}_{P} = \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{D}$

Q3. $P(\text{Spam}) = 0.2$ $P(\text{Not Spam}) = 0.8$

$P(\text{당첨} \mid \text{Spam}) = 0.5$

$P(\text{당첨} \mid \text{Not Spam}) = 0.01$

$P(\text{당첨}) = P(\text{당첨} \mid \text{Spam}) \cdot P(\text{Spam}) + P(\text{당첨} \mid \text{Not Spam}) \cdot P(\text{Not Spam})$
$= 0.5 \cdot 0.2 + 0.01 \cdot 0.8 = 0.1 + 0.008 = 0.108$

By Bayes rule, $P(\text{Spam} \mid \text{당첨}) = \dfrac{P(\text{당첨} \mid \text{Spam}) \cdot P(\text{Spam})}{P(\text{당첨})} = \dfrac{0.1}{0.108} = 0.9259.$

Q4. (2) 어떤 p에 대해 이항분포를 따르는 확률 변수의 엔트로피를 계산하기 위해서는 p를 알아야 한다.

하지만 p는 모수로 알 수 없으므로, p 대신 $\hat{p}$으로 p를 추정한다.

주어진 데이터에서 $\hat{p} = 9/15$ 이다.

entropy $H(\text{스윙 결과}) = H(\hat{p} = 9/15) = -9/15 \log_2(9/15) - 6/15 \log_2(6/15) = 0.9710$

(3) 두 명의 스윙 결과의 분포의 차이는 KL-divergence로 계산할 수 있다.

$D_{KL}(p = 9/15 \,\|\, q = 4/15) = 9/15 \log(9/4) + 6/15 \log(6/11) = 0.1655$

Q5. Convex Optimization of Logistic Regression

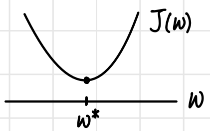
Loss function of Logistic Regression : Cross-Entropy Loss

Cross Entropy $H(y^{(i)}, \hat{y}^{(i)}) = -y^{(i)} \log \hat{y}^{(i)} - (1-y^{(i)}) \log(1-\hat{y}^{(i)})$

Goal : $\min_w \sum_{i=1}^{m} -y^{(i)} \log \hat{y}^{(i)} - (1-y^{(i)}) \log(1-\hat{y}^{(i)})$

$$= \min_w \sum_{i=1}^{m} \underbrace{-y^{(i)} \log \frac{1}{1+e^{-w'x^{(i)}}}}_{\text{Convex}} \underbrace{-(1-y^{(i)}) \log \frac{e^{-w'x^{(i)}}}{1+e^{-w'x^{(i)}}}}_{\text{Convex}} = \min_w \sum_{i=1}^{m} \underbrace{J(w)}_{\text{Convex}}$$

Since $J(w)$ is convex, we can find $w = w^*$ s.t. $\nabla J(w) = 0$.



This can be done via Gradient Descent method.