EE412 Foundation of Big Data Analytics, Fall 2021
HW3

Name: 노현섭

Student ID: 20190220

Discussion Group (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

Answer to Problem 1

Exercise 5.1.2

```
import numpy as np

M = np.array([1/3, 1/2, 0, 1/3, 0, 1/2, 1/3, 1/2, 1/2]).reshape((3,3))
M = np.identity(3) - 0.8*M
e = 0.2*np.array([1/3, 1/3, 1/3]).reshape((3,1))

print(np.dot(np.linalg.inv(M), e))
```

5.1.2

$$M = \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/3 & 1/2 \end{bmatrix}, \quad v' = 0.8 \begin{bmatrix} 1/3 & 1/2 & 0 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/3 & 1/2 \end{bmatrix} v + 0.2 \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \quad (\because \beta = 0.8)$$

$$\text{first } v = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Result = [[0.25925926]
[0.30864198]
[0.43209877]]

Exercise 5.3.1

```
import numpy as np

M = np.array([0, 1/2, 1, 0, 1/3, 0, 0, 1/2, 1/3, 0, 0, 1/2, 1/3, 1/2, 0, 0]).reshape((4,4))
M = np.identity(4) - 0.8*M
e = 0.2*np.array([1, 0, 0, 0]).reshape((4,1))       # For a: A only
#e = 0.2*np.array([1/2, 0, 1/2, 0]).reshape((4,1))  # For b: A and C
print(np.dot(np.linalg.inv(M), e))
```

5.3.1

$$M = \begin{bmatrix} 0 & 1/2 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}, \quad \beta = 0.8$$

(a) case : $S = \{A\}$, $v' = \beta M v + (1-\beta)\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, first $v = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

(b) case : $S = \{A, C\}$, $v' = \beta M v + (1-\beta)\begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$, first $v = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$

(a) A only = [[0.42857143]
[0.19047619]
[0.19047619]
[0.19047619]]

(b) A and C = [[0.38571429]
[0.17142857]
[0.27142857]
[0.17142857]]

Answer to Problem 2

Exercise 10.3.2

10.3.2

(a) $n=20,\ d=5 \longrightarrow 20 \times \binom{5}{t} / \binom{20}{t} \geq s$

$\therefore\ t=1 \rightarrow s=5$

maximal pairs : $(1,5)$

(b) $n=200,\ d=150 \longrightarrow 200 \times \binom{150}{t} / \binom{200}{t} \geq s$

$\therefore\ t=1 \rightarrow s=150,\ t=2 \rightarrow s=112,\ \cdots,\ t=10 \rightarrow s=11$

maximal pairs : $(1,150), (2,112), (3,84), (4,63), (5,49), (6,35), (7,26), (8,20), (9,15), (10,11)$

Exercise 10.5.2

10.5.2

(a)

$P_{wx} = 1-(1-P_C) = P_C,\quad P_{yz} = 1-(1-P_D) = P_D$

$P_{wy} = P_{wz} = P_{xy} = P_{xz} = \varepsilon$

$\therefore\ MLE = P_{wy}P_{wz}P_{xy}P_{xz}(1-P_{wz})(1-P_{xz}) = P_C P_D \varepsilon^2 (1-\varepsilon)^2$

for maximizing, it should be $P_C = P_D = 1$

therefore, when each members in the set C and D have edge among them (100%), it would be maximum likelihood with Figure 10.20

(b)

$P_{wx} = P_{wy} = P_{wz} = 1-(1-P_C) = P_C,$

$P_{xy} = P_{xz} = P_{yz} = 1-(1-P_C)(1-P_D) = P_C + P_D - P_C P_D$

$$\begin{aligned}\therefore\ MLE &= P_{wx}P_{wy}P_{xy}P_{yz}(1-P_{wz})(1-P_{xz}) \\ &= P_C^2 (P_C+P_D-P_CP_D)^2 (1-P_C)(1-P_C-P_D+P_CP_D) \\ &= P_C^2 (P_C+P_D-P_CP_D)^2 (1-P_C)^2 (1-P_D)\end{aligned}$$

for maximizing, it should be $P_C = \frac{2}{3},\ P_D = 0$

therefore, when 1) members in the set C have edge by the probability of $\frac{2}{3}$ and 2) members in the set D have edge by the probability of 0 (no have), it would be maximum likelihood with Figure 10.20

Answer to Problem 3

Exercise 12.5.3

12.5.3

(a)

GINI impurity (let $f(x)$) $\longrightarrow$ $f(x) = 1 - x^2 - (1-x)^2 = -2x^2 + 2x$

$x, y, z \in \text{dom} f$ s.t. $x < y < z$

$\frac{y-z}{y-x} f(x) + \frac{z-x}{y-x} f(y) = [(y-z)(-2x^2+2x) + (z-x)(-2y^2+2y)] / (y-x)$

$= (-2x^2y + 2xy + 2x^2z - 2xz - 2y^2z + 2yz + 2xy^2 - 2xy) / (y-x)$

$= [2xy(y-x) - 2z(y+x)(y-x) + 2z(y-x)] / (y-x)$

$= 2xy - 2z(x+y) + 2z$

$f(z) = -2z^2 + 2z$

$\rightarrow$ If it is concave, $2xy - 2xz - 2yz + 2z - f(z) < 0$

$\rightarrow$ $xy - xz - yz + z^2 < 0$

$\rightarrow$ $x(y-z) - z(y-z) < 0$

$\rightarrow$ $x - z < 0$ $(\because y > z)$

$\rightarrow$ $x < z$ $\therefore$ satisfy the assumption $(x < z < y)$ and condition.

$\therefore$ concave

(b)

Entropy measure of impurity (let $f(x)$) $\longrightarrow$ $f(x) = -x\log x - (1-x)\log(1-x)$

$(\text{dom} f \in [0,1])$

if we prove about convexity of $-f(x)$, then $f(x)$ is concave.

$\rightarrow$ $-f(x) = x\log x + (1-x)\log(1-x)$

by 2nd order condition of convexity, if $\nabla^2(-f(x)) \geq 0$ $\forall x \in \text{dom} f$,

$-f(x)$ is convex

$\rightarrow$ $\nabla(-f(x)) = \log x + \frac{1}{\ln 2} - \log(1-x) - \frac{1}{\ln 2} = \log x - \log(1-x)$

$\rightarrow$ $\nabla^2(-f(x)) = \frac{1}{x\ln 2} + \frac{1}{(1-x)\ln 2} \geq 0$ $(\because 0 \leq x \leq 1)$

$\therefore$ $-f(x)$ is convex and $f(x)$ is concave

$\rightarrow$ Entropy measure of impurity is concave