Theory Questions

1.

<a>

Compute the mean:

$$\mu_1 = \frac{-2-5-3+0-8-2+1+5-1+6}{10} = -0.9$$

$$\mu_2 = \frac{1-4+1+3+11+5+0-1-3+1}{10} = 1.4$$

Compute the Standard deviation:

$$\sigma_1^2 = \frac{1}{10-1}\times[|-2-\mu_1|^2 + |-5-\mu_1|^2 + |-3-\mu_1|^2 + |0-\mu_1|^2 + |-8-\mu_1|^2 + |-2-\mu_1|^2 + |1-\mu_1|^2 + |5-\mu_1|^2 + |-1-\mu_1|^2 + |6-\mu_1|^2] = \frac{1609}{90}$$

$$\sigma_1 = \sqrt{\frac{1609}{90}} \approx 4.2282$$

$$\sigma_2^2 = \frac{1}{10-1}\times[|1-\mu_2|^2 + |-4-\mu_2|^2 + |1-\mu_2|^2 + |3-\mu_2|^2 + |11-\mu_2|^2 + |5-\mu_2|^2 + |0-\mu_2|^2 + |-1-\mu_2|^2 + |-3-\mu_2|^2 + |1-\mu_2|^2] = \frac{274}{15}$$

$$\sigma_2 = \sqrt{\frac{274}{15}} \approx 4.2740$$

Standardize data:

$$data = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ +0.2129 & +0.3744 \\ -1.6793 & +2.2462 \\ -0.2602 & +0.8423 \\ +0.4494 & -0.3276 \\ +1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ +1.6319 & -0.0936 \end{bmatrix}$$

Find covariance matrix:

$$cov = data^T \times data \div (10-1) = \begin{bmatrix} 9 & -3.6744 \\ -3.6744 & 9 \end{bmatrix} \div 9 = \begin{bmatrix} 1 & -0.4083 \\ -0.4083 & 1 \end{bmatrix}$$

Find eigenvalues:

$$\begin{bmatrix} 1 & -0.4083 \\ -0.4083 & 1 \end{bmatrix} - \lambda I = 0$$

$$(1-\lambda)^2 - (-0.4083)^2 = 0 \Rightarrow \lambda_1 = 0.5917, \lambda_2 = 1.4083$$

$$\lambda = \begin{bmatrix} 0.5917 \\ 1.4083 \end{bmatrix}$$

Find eigenvectors:

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

for $\lambda = 0.5917$:

$$\begin{bmatrix} 1 - 0.5917 & -0.4083 \\ -0.4083 & 1 - 0.5917 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$V_1 = \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}$$

for $\lambda = 1.4083$:

$$\begin{bmatrix} 1 - 1.4083 & -0.4083 \\ -0.4083 & 1 - 1.4083 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$V_2 = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}$$

<b>

Project the data:

$$data \times V_2 = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ +0.2129 & +0.3744 \\ -1.6793 & +2.2462 \\ -0.2602 & +0.8423 \\ +0.4494 & -0.3276 \\ +1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ +1.6319 & -0.0936 \end{bmatrix} \times \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix} = \begin{bmatrix} 0.1178 \\ -0.2077 \\ 0.2850 \\ 0.1142 \\ 2.7756 \\ 0.7796 \\ -0.5494 \\ -1.3838 \\ -0.7112 \\ -1.2201 \end{bmatrix}$$

2.
<a>

$$H\left(P(v_1), \dots, P(v_n)\right) = \sum(-P(v_i)\log P(v_i))$$
$$remainder(A) = \sum \frac{p_i + n_i}{p + n} H(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i})$$

Compute for feature 1:

$$remainder(1) = \frac{1+1}{5+5} H\left(\frac{1}{1+1}, \frac{1}{1+1}\right) + 4\times\frac{1+0}{5+5} H\left(\frac{1}{1+0}, \frac{0}{1+0}\right) + 4\times\frac{0+1}{5+5} H\left(\frac{0}{0+1}, \frac{1}{0+1}\right) = 0.2$$

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$$

$$H(1,0) = -1\log 1 - 0\log 0 = 0$$

$$IG(1) = 1 - 0.2 = 0.8$$

Compute for feature 2:

$$remainder(2) = \frac{2+1}{5+5} H\left(\frac{2}{2+1}, \frac{1}{1+2}\right) + 8\times\frac{1+0}{5+5} H\left(\frac{1}{1+0}, \frac{0}{1+0}\right) = 0.2755$$

$$H\left(\frac{2}{3}, \frac{1}{3}\right) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.9183$$

$$IG(2) = 1 - 0.2755 = 0.7245$$

<b>
Compare two features:

$$IG(1) > IG(2)$$

we should prioritize feature 1.

<c>
Standardize all the data:

Using the same computing method in previous question, find out
$\mu_1 = -0.9$, $\mu_2 = 1.4$, $\sigma_1 = 4.2282$, $\sigma_2 = 4.2740$
Then, the data becomes

$$C_1 = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ +0.2129 & +0.3744 \\ -1.6792 & +2.2462 \end{bmatrix} \text{ and } C_2 = \begin{bmatrix} -0.2602 & +0.8423 \\ +0.4494 & -0.3276 \\ +1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ +1.6319 & -0.0936 \end{bmatrix}$$

Compute the mean for each class:

$$\mu_1 = [-0.6386 \quad 0.2340], \quad \mu_2 = [0.6386 \quad -0.2340]$$

Compute scatter matrices for each class:

Using formula $\sigma^2 = (N-1)\times cov(C)$, where $cov(C) = C^T C \div (N-1)$

$$\sigma_1^2 = (5-1)\times cov(C_1) = \begin{bmatrix} 2.0808 & -1.6490 \\ -1.6490 & 6.5255 \end{bmatrix}$$

$$\sigma_2^2 = (5-1)\times cov(C_2) = \begin{bmatrix} 2.8415 & -0.5312 \\ -0.5312 & 1.9270 \end{bmatrix}$$

Within class scatter matrix:

$$S_W = \sigma_1^2 + \sigma_2^2 = \begin{bmatrix} 4.9223 & -2.1803 \\ -2.1803 & 8.4526 \end{bmatrix}$$

$S_W^{-1} = \begin{bmatrix} 0.2294 & 0.0592 \\ 0.0592 & 0.1336 \end{bmatrix}$, by using row reduction

Preform eigen-decomposition:

$$S_B = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = \begin{bmatrix} 1.6311 & -0.5976 \\ -0.5976 & 0.2190 \end{bmatrix}$$

$$S_W^{-1} S_B = \begin{bmatrix} 0.2294 & 0.0592 \\ 0.0592 & 0.1336 \end{bmatrix} \begin{bmatrix} 1.6311 & -0.5976 \\ -0.5976 & 0.2190 \end{bmatrix} = \begin{bmatrix} 0.3388 & -0.1241 \\ 0.0167 & -0.0061 \end{bmatrix}$$

Eigen-values and eigen-vector:

Compute for eigen-values, got the result $\lambda = \begin{bmatrix} 0.3327 \\ 0 \end{bmatrix}$.

Compute eigen-vector by using non-zero eigen-value 0.3327: $W = \begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix}$.

<d>

Project the data of Class 1:

$$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ +0.2129 & +0.3744 \\ -1.6792 & +2.2462 \end{bmatrix} \begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix} = \begin{bmatrix} -0.2645 \\ -1.0308 \\ -0.5007 \\ 0.2311 \\ -1.5664 \end{bmatrix}$$

Project the data of Class 2:

$$\begin{bmatrix} -0.2602 & +0.8423 \\ +0.4494 & -0.3276 \\ +1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ +1.6319 & -0.0936 \end{bmatrix} \begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix} = \begin{bmatrix} -0.2184 \\ 0.4327 \\ 1.3660 \\ -0.0744 \\ 1.6253 \end{bmatrix}$$

<e>

Consider the projected data as feature 3 and compute the information gain for that feature.

$remainder(3) = 10\times\frac{1+0}{5+5}H\left(\frac{1}{1+0}, \frac{0}{1+0}\right) = 0$, $IG(3) = 1 - 0 = 1$, which is greater than $IG(1)$.

For this feature, most data in class 1 are smaller and most data in class 2 are larger.

I conclude that the projection I performed seem to provide a good class separation.

Dimensionality Reduction via PCA