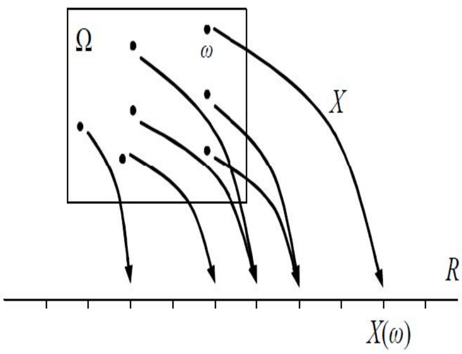
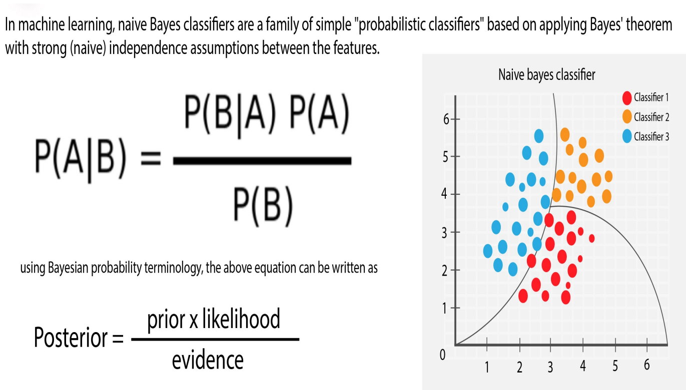
Αναγνώριση Προτύπων

2021-2022

Εργασία 2η

ΠΙΘΑΝΟΤΗΤΕΣ-ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ, BAYESIAN ΘΕΩΡΙΑ ΑΠΟΦΑΣΕΩΝ





Καρυπίδης Ευστάθιος 57556

12/11/21, Ξάνθη

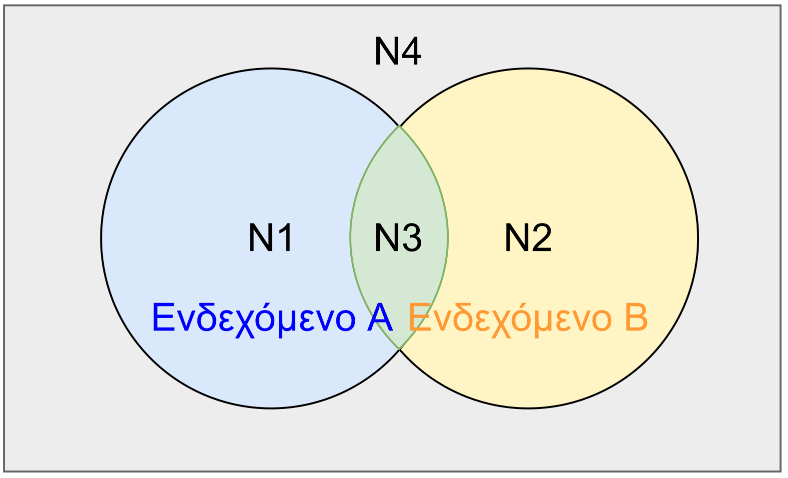
Για την υλοποίηση των ερωτημάτων της εργασίας έγινε χρήση γλώσσας python και των βιβλιοθηκών numpy,scipy και matplotlib.

Άσκηση 1

Έχουμε ενα πείραμα τύχης και τα ενδεχόμενα A, B. Δίνονται τα παρακάτω δεδομένα:

- Μόνο το A συνέβη N1 φορές
- Μόνο το B συνέβη N2 φορές
- Το A και B συνέβη N3 φορές
- Ούτε το A ούτε το B συνέβη N4 φορές

Για να είναι πιο κατανοητό μπορούν τα δεδομένα να απεικονιστούν σε διάγραμμα Venn.



Αρχικά, ορίζουμε πως ο συνολικός αριθμός πειραμάτων είναι:

$$N = N_1 + N_2 + N_3 + N_4$$

Για τη δεσμευμένη πιθανότητα(πιθανότητα του A δεδομένου ότι έχει συμβεί το B) γνωρίζουμε:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Απο εκφώνηση γνωρίζουμε πως το A, B συνέβει N_3 φορές άρα η πιθανότητα $P(A \cap B)$ είναι N_3 προς τον συνολικό αριθμό πειραμάτων, δηλαδή:

$$P(A \cap B) = \frac{N_3}{N}$$

Στη συνέχεια χρειάζεται να υπολογιστεί η πιθανότητα του B. Στη περίπτωση αυτή πρέπει να βρεθούν πόσες φορές συνέβει το B. Από τη θεωρία συνόλων γνωρίζουμε ότι:

Εμφανίσεις του B = Εμφανίσεις μονο του B(B-A) + Εμφανίσεις του A και B ($A \cap B$) = $N_2 + N_3$ δηλαδή:

$$P(B) = \frac{N_2 + N_3}{N}$$

Άρα τελικά έχουμε ότι:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{N_3}{N}}{\frac{N_2+N_3}{N}} = \frac{N_3}{N_2 + N_3}$$

Άσκηση 2

Δίνονται οι τιμές τις υπο συνθήκη πιθανότητες μιας τυχαίας μεταβλητής X για 3 κατηγορίες $\omega_1, \omega_2, \omega_3$ (τρία ζάρια) η αλλιώς η συνάρτηση πιθανοφάνειας (likelihood function) $P(X = x|\omega)$

p(X=x|ω)						
	X=1	X=2	X=3	X=4	X=5	X=6
ω_1	0.3	0.2	0.1	0.1	0.2	0.1
ω_2	0.2	0.2	0.4	0.05	0.1	0.05
ω_3	0.1	0.3	0.15	0.05	0.3	0.1

Ακόμη δίνονται οι τιμές για τις (εκ των προτέρων) a priori πιθανότητες:

- $p(\omega_1) = 0.3$
- $p(\omega_2) = 0.3$

Άρα, η εκ των προτέρων πιθανοτητα για το ζάρι κατηγορίας ω_3 θα είναι:

$$p(\omega_3) = 1 - p(\omega_1) - p(\omega_2) = 1 - 0.3 - 0.3 \Rightarrow$$
$$p(\omega_3) = 0.4$$

Εφόσον δεν δίνεται τό κόστος σε περίπτωση λάθους απόφασης θεωρούμε πώς όλα τα κόστα ταξινόμησης είναι 1 δηλαδή:

$$\lambda\left(\alpha_j \mid \omega_i\right) = \begin{cases} 0 & j = i \\ 1 & j \neq i \end{cases} \quad j, i = 1, \ldots, c$$

όπου a_j η απόφαση ταξινόμησης και ω_i η πραγματική κλάση.

Σύμφωνα με τη θεωρία κανόνας ο απόφασης Bayes για ελαχιστοποίηση του ρίσκου υποδεικνύει την απόφαση η οποία ελαχιστοποιεί το υπό-συνθήκη ρίσκο. Δηλαδή, για να επιλεχθεί ως απόφαση η κλάση ω_i μεταξύ k κλάσεων πρέπει να ισχύει για την εκ των υστέρων πιθανότητα ότι:

$$p(\omega_i|x) > p(\omega_j|x) \text{ για όλα } j \neq i$$

Όμως ισχυεί ότι

$$p(\omega_i \cap x) = p(\omega_i|x) \cdot p(x) = p(\omega_i)p(x|\omega_i)$$

Ακόμη παρατηρούμε πώς ισχυεί ότι:

$$\sum_x p(x|\omega_i) = 1 \text{ για όλα τα } \omega_i \text{ δηλαδή για } i = 1, 2, 3$$
$$\text{άρα μπορούμε να πούμε ότι για το evidence } p(x):$$
$$p(x) = \sum_i p(x|\omega_i)p(\omega_i) = \sum_i p(\omega_i) = 1$$

Άρα από τον κανόνα/κριτήριο απόφασης Bayes (στη γενική περίπτωση για περισσότερες απο 2 κατηγορίες) για να επιλέξουμε μια κλάση/κατηγορία, δηλαδή στο πρόβλημα μας ένα ζάρι ω_i μεταξύ 3 κατηγοριών πρέπει να ισχυεί ότι :

$$p(\omega_i)p(x|\omega_i) > p(\omega_j)p(x|\omega_j)\ \forall j \in \{1, 2, 3\} \text{ και } j \neq i$$
$$\text{η αλλιώς}$$
$$p(\omega_i)p(x|\omega_i) = max[p(\omega_1)p(x|\omega_1), p(\omega_2)p(x|\omega_2), p(\omega_3)p(x|\omega_3)]$$

Μέσω του παρακάτω κώδικα python υπολογίζω τα γινόμενα.

```python
import numpy as np
likelihood = np.array([[0.3, 0.2, 0.1, 0.1, 0.2, 0.1],
                       [0.2, 0.2, 0.4, 0.05, 0.1, 0.05],
                       [0.1, 0.3, 0.15, 0.05, 0.3, 0.1]])
a_priori = np.array([0.3, 0.3, 0.4])
prod = np.zeros((3, 6))
for i in range(likelihood.shape[1]):
    prod[:, i] = np.multiply(likelihood[:, i], a_priori)
print(prod)
```

Το αποτέλεσμα φαίνεται παρακάτω. Σε κόκκινο πλαίσιο τοποθετούνται τα μέγιστα.

```
[[0.09  0.06  0.03  0.03  0.06  0.03 ]
 [0.06  0.06  0.12  0.015 0.03  0.015]
 [0.04  0.12  0.06  0.02  0.12  0.04 ]]
```

```
[[0.09  0.06  0.03  0.03  0.06  0.03 ]
 [0.06  0.06  0.12  0.015 0.03  0.015]
 [0.04  0.12  0.06  0.02  0.12  0.04 ]]
```

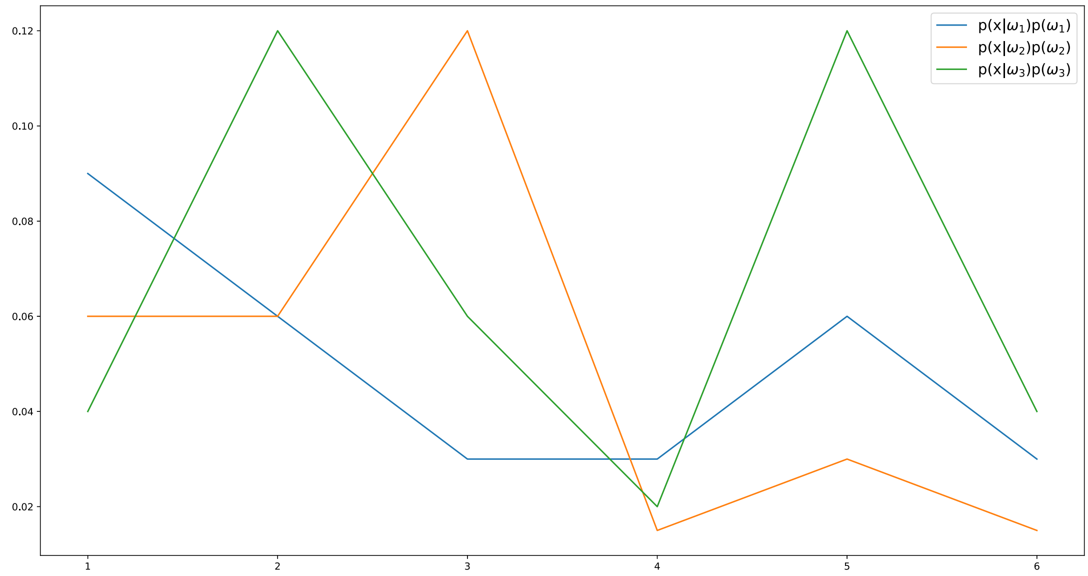
Με βάση κριτήριο απόφασης το ολικό ρίσκο (σφάλμα) ταξινόμησης(εάν ω_i η σωστή απόφαση) θα είναι ίσο με:

$$R_{total} = \sum_{i \neq j} p(\omega_j)p(x|\omega_j) = 1 - \sum_k p(\omega_i)p(x_k|\omega_i)$$
$$R_{total} = 1 - (0.09 + 0.12 + 0.12 + 0.03 + 0.12 + 0.04) = 1 - 0.52 = 0.48$$

Ακόμη με το παρακάτω κώδικα μπορούμε να το επαληθεύσουμε αλλά και να κάνουμε plot όπου φαίνονται και οι αποφάσεις ταξινόμησης(το μέγιστο σε κάθε περίπτωση)

```python
print("Total risk = ", 1 - np.sum(np.max(prod, axis = 0)))
for j in range(prod.shape[0]):
    plt.plot(prod[j, :])
plt.xticks(np.arange(prod.shape[1]), np.arange(1, prod.shape[1]+1))
plt.legend(['p(x|$\omega_1$)p($\omega_1$)', 'p(x|$\omega_2$)p($\omega_2$)', 'p(x|$\omega_3$)p($\omega_3$)'] prop={'size':
15})
plt.show()
```

```
Total risk =  0.48
```



Άσκηση 3

Δίνονται 2 κατηγορίες ω_1, ω_2, και πιθανότητα $P(\omega_1) = 0.25$. Συνεπώς $P(\omega_2) = 1 - P(\omega_1) = 1 - 0.25 = 0.75$. Επιπλέον, δίνεται ότι τα δείγματα προέρχονται από Γκαουσιανές κατανομές που περιγράφονται από τις ακόλουθες πυκνότητες πιθανότητας:

$$p\left(\bar{x}/\omega_1\right) = N\left(\bar{\mu}_1, \bar{\Sigma}_1\right) \text{ και } p\left(\bar{x}/\omega_2\right) = N\left(\bar{\mu}_2, \bar{\Sigma}_2\right)$$
$$\bar{\mu}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix} \bar{\mu}_2 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix}$$

Για την εύρεση της εξίσωσης επιφάνειας απόφασης και του ολικού σφάλματος κατηγοριοποίησης κατα Bayes χρησιμποιήθηκαν πληροφορίες πέραν των διαφανιών, από το βιβλίο του Duda, Pattern Classification 2ond Edition και ειδκότερα παράγραφοι 2.5 και 2.6.

Παρατηρούμε ότι στη συγκεκριμένη περίπτωση οι πινακες συνδιασποράς Σ είναι ίσοι δηλαδή, $\Sigma = \Sigma_1 = \Sigma_2$. Αυτή η περίπτωση αποτελεί ειδική περίπτωση όπου εξετάζεται στη παράγραφο 2.6.2 και οι συναρτήσεις διάκρισης είναι γραμμικές και ίσες με:

$$g_i(x) = w_i^T x + \omega_{i0} \text{ όπου}$$
$$w_i = \Sigma^{-1}\mu_i \text{ και } \omega_{i0} = -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \ln P\left(\omega_i\right)$$

Αρχικά, υπολογίζουμε τον αντίστροφο πίνακα διασποράς:

$$\Sigma^{-1} = \frac{1}{det(\Sigma)}\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix}$$

Ακόμη ισχυεί ότι:

$$\left(\Sigma^{-1}\right)^T = \Sigma^{-1} = \frac{1}{35}\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix}$$

Για τις επιφάνειες απόφασης έχουμε ότι:

$$
\begin{gathered}
g_1(x)=g_2(x) \Rightarrow \\
g_1(x)-g_2(x)=0 \Rightarrow \\
\left(\Sigma^{-1} \mu_1\right)^T \mathrm{x}-\frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1-\left(\Sigma^{-1} \mu_2\right)^T \mathrm{x}+\frac{1}{2} \mu_2^t \Sigma^{-1} \mu_2+\ln \left(\frac{P\left(\omega_1\right)}{P\left(\omega_2\right)}\right)=0 \Rightarrow
\end{gathered}
$$

Υπολογίζω ξεχωριστά το $w^T x$:

$$
\begin{gathered}
w^T x=(\Sigma^{-1} \mu_1)^T x-(\Sigma^{-1} \mu_2)^T x \Rightarrow \\
w^T x=\frac{1}{35}(\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix})^T x-\frac{1}{35}(\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} -2 \\ -1 \end{bmatrix})^T x \Rightarrow \\
w^T x=\frac{1}{35}[7 \quad 7] x-\frac{1}{35}[-17 \quad -2] x \Rightarrow \\
w^T x=\frac{1}{35}[24 \quad 9] x \Rightarrow \\
w=\frac{1}{35}\begin{bmatrix} 24 \\ 9 \end{bmatrix}
\end{gathered}
$$

Αντίστοιχα για το w_0

$$
\begin{gathered}
w_0=-\frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1+\frac{1}{2} \mu_2^t \Sigma^{-1} \mu_2+\ln \left(\frac{P(\omega_1)}{P(\omega_2)}\right) \Rightarrow \\
w_0=-\frac{1}{2} \frac{1}{35} 21+\frac{1}{2} \frac{1}{35} 36+\ln \left(\frac{0.25}{0.75}\right) \Rightarrow \\
w_0=\frac{1}{2} \frac{1}{35} 15+\ln \left(\frac{0.25}{0.75}\right)=3 / 14-\frac{713}{769} \Rightarrow \\
w_0=-0.884326574382395
\end{gathered}
$$

Άρα τελικά η επιφάνεια απόφασης για $x=\left[x_1 \quad x2\right]^T$ είναι $w^T x+w_0=0$:

$$
\begin{gathered}
\frac{24}{35} x_1+\frac{9}{35} x_2-0.884326574382395=0 \Rightarrow \text{Για προσέγγιση σε 4ο δεκαδικό} \\
\frac{24}{35} x_1+\frac{9}{35} x_2-0.8843=0
\end{gathered}
$$

Όσον αφορά το ολικό σφάλμα της ταξινόμησης, αυτό σύμφωνα με τη παράγραφο 2.7(Error Probabilities and Integrals) είναι:

$$
\begin{aligned}
P(\text { error }) & =P\left(\mathbf{x} \in \mathcal{R}_2, \omega_1\right)+P\left(\mathbf{x} \in \mathcal{R}_1, \omega_2\right) \\
& =P\left(\mathbf{x} \in \mathcal{R}_2 \mid \omega_1\right) P\left(\omega_1\right)+P\left(\mathbf{x} \in \mathcal{R}_1 \mid \omega_2\right) P\left(\omega_2\right) \\
& =\int_{\mathcal{R}_2} p\left(\mathbf{x} \mid \omega_1\right) P\left(\omega_1\right) d \mathbf{x}+\int_{\mathcal{R}_1} p\left(\mathbf{x} \mid \omega_2\right) P\left(\omega_2\right) d \mathbf{x} .
\end{aligned}
$$

Αυτός ο τύπος ερμηνέυεται ως εξής. Το σφάλμα από κάθε κατηγοριοποίηση είναι το άθροισμα της πιθανότητας να ανήκει το δείγμα στην κατηγορία ω_1 ενώ το ταξινομούμε στην κατηγορία ω_2 και της πιθανότητα σφάλματος το δείγμα να ανήκει στην κατηγορία ω_2 ενώ το ταξινομούμε στην κατηγορία ω_1. Όμως, εφόσον τα δείγματα είναι σε 2 διαστάσεις έτσι και τα ολοκληρώματα θα είναι διπλά. Αυτό καθιστά εξαιρετικά δύσκολη την επίλυση τους. Έπειτα από αναζήτηση, στο βιβλίο του Duda παράγραφο 2.5.2(Multivariate Density) σελίδα 19(του βιβλίου) και στο βιβλίο Στατιστική Αναγνώριση Προτύπων(Θεόδωρος Αλεξόπουλος-Αικατερίνη Τζαμαριουδάκη) στη σελιδα 55(64 του pdf) από τη βιβλιογραφία του eclass παρουσιάζεται η παρακάτω μέθοδος(πάνω απο Duda, κάτω απο Αλεξόπουλο):

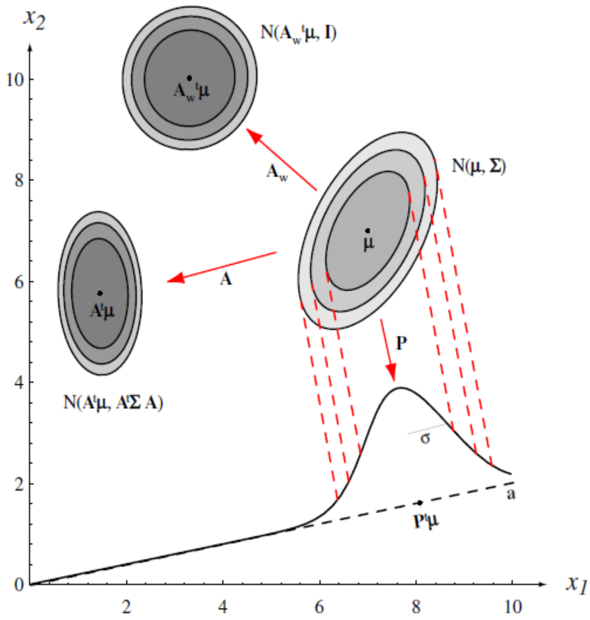
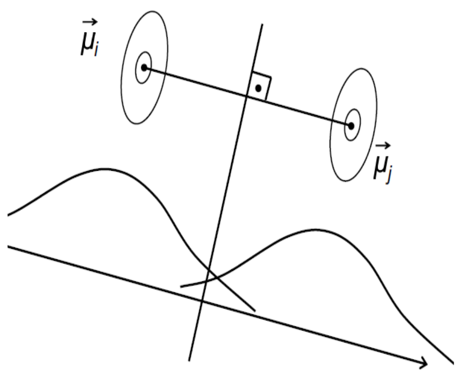


Figure 2.8: The action of a linear transformation on the feature space will convert an arbitrary normal distribution into another normal distribution. One transformation, **A**, takes the source distribution into distribution $N(\mathbf{A}^t\mu, \mathbf{A}^t\Sigma\mathbf{A})$. Another linear transformation — a projection **P** onto line **a** — leads to $N(\mu, \sigma^2)$ measured along **a**. While the transforms yield distributions in a different space, we show them superimposed on the original $x_1 - x_2$ space. A whitening transform leads to a circularly symmetric Gaussian, here shown displaced.

Η ανάλυση σφάλματος είναι πολύ δύσκολη για αυτό το παράδειγμα. Ένας τρόπος είναι το ακόλουθο όχι και τόσο πραγματικό παράδειγμα:



Σχήμα 2.21

Σε αυτή την περίπτωση κάνουμε προβολή σε μία διάσταση και κάνουμε ανάλυση σφάλματος σε μία διάσταση. Αν $\mathbf{x}$ είναι γκαουσιανή τότε το $y = \mathbf{w}^t_{1\times n}\mathbf{x}_{n\times 1}$ είναι γκαουσιανή σε μία διάσταση

$$y \sim N\left(\mathbf{w}^t_{1\times n}\mu,\ \mathbf{w}^t_{1\times n}\Sigma_{n\times n}\mathbf{w}_{n\times 1}\right)$$

Ουσιαστικά θέτουμε όπου $y = w^T x$ και έτσι απο 2 διαστάσεις του x καταλήξαμε στο y όπου είναι γκαουσιανή σε μία διάσταση. Άρα, το σημείο τομής θα είναι $y - 0.8843 = 0 \Rightarrow y = 0.8843$. Δηλαδή έχουμε ω_1 όταν $y > 0.8843$ και ω_2 όταν $y < 0.8843$. Άρα θα έχουμε ότι:

$$p\left(x \mid \omega_i\right) = \mathrm{N}\left(\mu_i, \Sigma_i\right) => p\left(w^T x \mid \omega_i\right) = \mathrm{N}\left(w^T \mu_i, w^T \Sigma w\right)$$

$$\text{Από πρίν έχουμε ότι } w = \frac{1}{35}\begin{bmatrix} 24 \\ 9 \end{bmatrix}$$

$$w^T \mu_1 = \begin{bmatrix} \frac{24}{35} \\ \frac{9}{35} \end{bmatrix}^T \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{24}{35} & \frac{9}{35} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{6}{5} = 1.2$$

$$w^T \mu_2 = \begin{bmatrix} \frac{24}{35} \\ \frac{9}{35} \end{bmatrix}^T \begin{bmatrix} -2 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{24}{35} & \frac{9}{35} \end{bmatrix} \begin{bmatrix} -2 \\ -1 \end{bmatrix} = -\frac{57}{35} = -1.628571428571429 \approx -1.6286$$

$$w^T \Sigma_1 w = \begin{bmatrix} \frac{24}{35} \frac{9}{35} \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \frac{24}{35} \\ \frac{9}{35} \end{bmatrix} = 99/35 = 2.828571428571428 \approx 2.8286$$

$$w^T \Sigma_2 w = \begin{bmatrix} \frac{24}{35} & \frac{9}{35} \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix} \begin{bmatrix} \frac{24}{35} \\ \frac{9}{35} \end{bmatrix} = 99/35 = 2.828571428571428 \approx 2.8286$$

Συνεπώς αντικαθιστώντας έχουμε:

$$P_e = P\left(\omega_1\right) \int_{-\infty}^{0.8843} \mathrm{N}(1.2, \sqrt{2.8286}) dx + P\left(\omega_2\right) \int_{0.8843}^{\infty} \mathrm{N}(-1.6286, \sqrt{2.8286}) dx$$

Θέλει ιδιαίτερη προσοχή καθώς στα άκρα ολοκήρωσης στο 1ο ολοκλήρωμα έχουμε την περιοχή όπου αναφέρεται στο σφάλμα ταξινόμησης των ω_1 ως ω_2(στην περιοχή R_2) και στο 2ο ολοκλήρωμα το αντίστροφο εφόσον έχουμε ω_1 όταν $y > 0.8843$ και ω_2 όταν $y < 0.8843$.

Τώρα ομως γνωρίζουμε ότι το ολοκληρώματα αυτά συνδέονται με την erf() και την Normal CDF:

$$P_e = P\left(\omega_1\right) \Phi\left(\frac{0.8843 - 1.2}{\sqrt{2.8286}}\right) + P\left(\omega_2\right)\left(1 - \Phi\left(\frac{0.8843 + 1.6286}{\sqrt{2.8286}}\right)\right)$$

όπου $\Phi(x)$ η Standart Normal CDF η οποία μπορεί να υπολογιστεί απο το matlab με την εντολή normcdf(x):

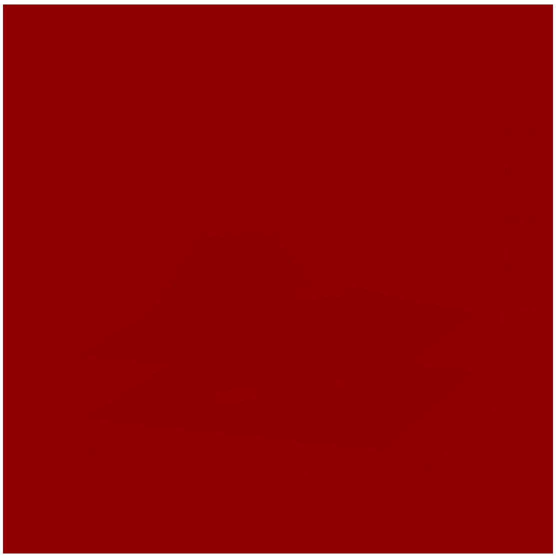
$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

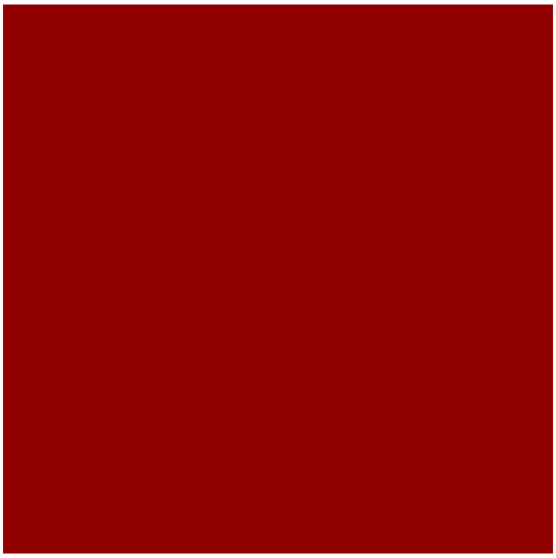
```
Pe = (1/4)*normcdf((0.8843-1.2)/sqrt(2.8286))+(3/4)-(3/4)*normcdf((0.8843+1.6286)/sqrt(2.8286))
```

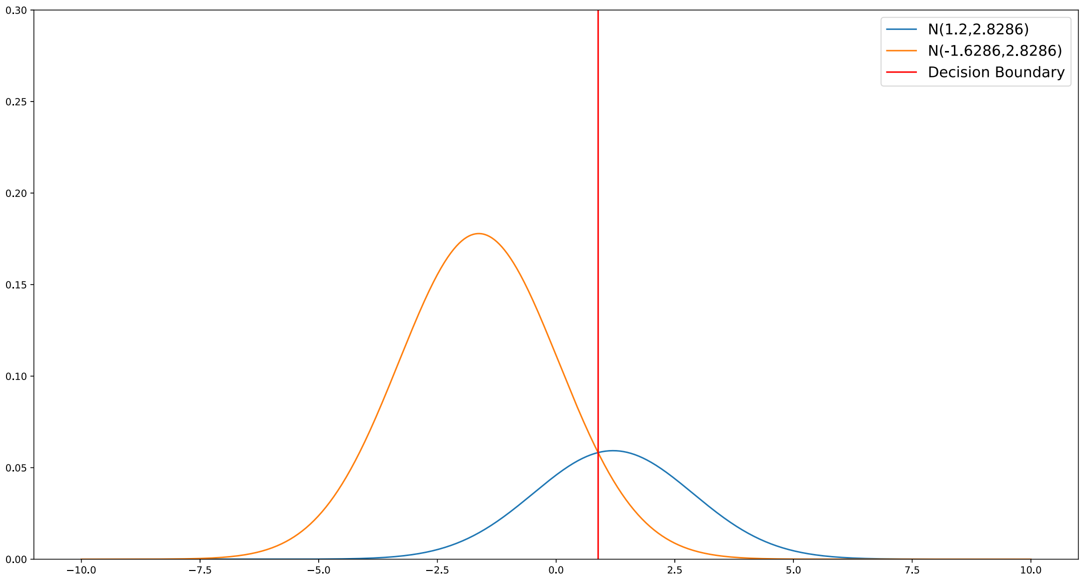
```
Pe =

   0.157065701297434
```

Στο αρχείο HW2_Exercise3_4_visualizations.py περιέχεται ο κώδικας για την οπτικοποίηση των αποτελεσμάτων. Για τη συγκεκριμένη άσκηση μπορούμε να δούμε την επιφάνεια απόφασης. Όπως βλέπουμε και από τα θεωρητικάη μία επιφάνεια






0.30
0.25
0.20
0.15
0.10
0.05
0.00
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
N(1.2,2.8286)
N(-1.6286,2.8286)
Decision Boundary

Άσκηση 4

Ερώτημα α

Έστω ένα πρόβλημα ταξινόμησης 2 κλάσεων $\Omega = \{\omega_1, \omega_2\}$ όπου δίνονται οι πιθανοφάνειες $P(x|\omega_1) = N(2, 0.5)$ και $P(x|\omega_2) = N(1.5, 0.2)$ οι a-priori πιθανότητες $P(\omega_1) = \frac{1}{3}$, $P(\omega_1) = \frac{2}{3}$ και $\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$.

Για να βρούμε τη βέλτιστη λύση ξέρουμε από τη θεωρία ότι για να κατηγοριοποιήσουμε κάτι στην κλάση $\omega 1$ πρέπει να ισχυεί:

$$\text{Για } \lambda_{21} > \lambda_{11} = 3 > 1 \text{ Αληθές}$$
$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} \geq \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

Αντικαθιστώντας έχουμε:

$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} \geq \frac{2-1}{3-1} \frac{\frac{2}{3}}{\frac{1}{3}}$$
$$\frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} \geq 1$$

Στη συνέχεια αντικαθιστώ και τις a-priori πιθανότητες:

$$\frac{\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}{\frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}} \geq 1 \Rightarrow$$
$$\frac{\sigma_2}{\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}} \geq 1 \Rightarrow$$
$$\ln\left(\frac{\sigma_2}{\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) \geq \ln(1) \Rightarrow$$
$$\ln\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2} \geq 0 \Rightarrow$$
$$\ln\left(\frac{\sqrt{0.2}}{\sqrt{0.5}}\right) + \frac{(x-1.5)^2}{2 \cdot 0.2} - \frac{(x-2)^2}{2 \cdot 0.5} \geq 0 \Rightarrow$$
$$\frac{(x-2)^2}{1} - \frac{(x-1.5)^2}{0.4} \leq -0.458145365937077 \Rightarrow$$
$$x^2 - 4x + 4 - 2.5x^2 + 7.5x - 5.625 \leq -0.458145365937077 \Rightarrow$$
$$1.5x^2 - 3.5x + 1.166854634062923 \geq 0$$

Οι ρίζες είναι $x_1 = 0.402986090944180,\ x_2 = 1.930347242389154$. Επειδή, το α>0, σημαίνει πως ανάμεσα στις ρίζες η συνάρτηση είναι αρνητική ενώ εκτός δηλαδή στα υπόλοιπα διαστήματα θετική. Αυτό, μπορούμε να το δούμε και εύκολα παίρνοντας τη μορφή με τις ρίζες δηλαδή,
$(x - 0.402986090944180)(x - 1.930347242389154) \geq 0$

Έτσι η ταξινόμηση γίνεται ως εξής:

$$x \in (-\infty, 0.402986090944180) \cup (1.930347242389154, +\infty) \text{ Ταξινομείται στην } \omega_1$$
$$x \in (0.402986090944180, 1.930347242389154) \text{ Ταξινομείται στην } \omega_2$$

Για το ελάχιστο ρίσκο γνωρίζουμε ότι:

$$C = \int_{\mathcal{R}_1} R(a_1|x)p(x)d\mathbf{x} + \int_{\mathcal{R}_2} R(a_2|x)p(x)d\mathbf{x}$$

Αντικαθιστώντας όπου:

$$R(a_1 \mid \mathbf{x}) = \lambda_{11}P(\omega_1 \mid \mathbf{x}) + \lambda_{12}P(\omega_2 \mid \mathbf{x})$$
$$R(a_2 \mid \mathbf{x}) = \lambda_{21}P(\omega_1 \mid \mathbf{x}) + \lambda_{22}P(\omega_2 \mid \mathbf{x})$$

και με χρήση του κανόνα του Bayes για τη δεσμευμένη πιθανότητα καταλήγουμε ότι:

$$C = \int_{R_1} [\lambda_{11}P(\omega_1)p(x \mid \omega_1) + \lambda_{12}P(\omega_2)p(x \mid \omega_2)]d\mathbf{x}$$
$$+ \int_{R_2} [\lambda_{21}P(\omega_1)p(x \mid \omega_1) + \lambda_{22}P(\omega_2)p(x \mid \omega_2)]d\mathbf{x} \Rightarrow$$
$$C = P(\omega_1)\left[\lambda_{11}\int_{R_1} p(x \mid \omega_1)d\mathbf{x} + \lambda_{21}\int_{R_2} p(x \mid \omega_1)d\mathbf{x}\right] +$$
$$P(\omega_2)\left[\lambda_{12}\int_{R_1} p(x \mid \omega_2)d\mathbf{x} + \lambda_{22}\int_{R_2} p(x \mid \omega_2)d\mathbf{x}\right]$$
$$C = \frac{1}{3}\left[1\int_{R_1} p(x \mid \omega_1)d\mathbf{x} + 3\int_{R_2} p(x \mid \omega_1)d\mathbf{x}\right] +$$
$$\frac{2}{3}\left[2\int_{R_1} p(x \mid \omega_2)d\mathbf{x} + 1\int_{R_2} p(x \mid \omega_2)d\mathbf{x}\right]$$

Ο συγκεκριμένος τύπος υπάρχει και στο βιβλίο του Duda, στη σελίδα 11, παράγραφο 2.3.1 ως σχέση (23). Στη συνέχεια χρησιμοποιώντας τις παρακάτω ιδιότητες και με χρήση της normcdf του μάτλαμπ καταλήγουμε στο ότι:

If X is a normal random variable with mean μ and variance σ^2, i.e, $X \sim N(\mu, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

$$F_X(x) = P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$P(a < X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

$$\int_{R_2} p\left(x \mid \omega_1\right) d\mathbf{x} = P\left([0.403 < x < 1.9303] \mid \omega_1\right) = \Phi\left(\frac{1.9303-2}{\sqrt{0.5}}\right) - \Phi\left(\frac{0.403-2}{\sqrt{0.5}}\right) \Rightarrow$$

$$\int_{R_2} p\left(x \mid \omega_1\right) d\mathbf{x} = 0.46074 - 0.01195 = 0.44879$$

$$\int_{R_2} p\left(x \mid \omega_2\right) d\mathbf{x} = P\left([0.403 < x < 1.9303] \mid \omega_2\right) = \Phi\left(\frac{1.9303-1.5}{\sqrt{0.2}}\right) - \Phi\left(\frac{0.403-1.5}{\sqrt{0.2}}\right) \Rightarrow$$

$$\int_{R_2} p\left(x \mid \omega_2\right) d\mathbf{x} = 0.83202 - 0.00708 = 0.82494$$

$$\int_{R_1} p\left(x \mid \omega_1\right) d\mathbf{x} = P([x < 0.403 \cup 1.93033 < x]) \mid \omega_1) = 1 - 0.44879 = 0.55121$$

$$\int_{R_1} p\left(x \mid \omega_2\right) d\mathbf{x} = P([x < 0.403 \cup 1.93033 < x]) \mid \omega_2) = 1 - 0.82494 = 0.17506$$

$$C = \frac{1}{3}(1 \cdot 0.55121 + 3 \cdot 0.44879) + \frac{2}{3}(2 \cdot 0.17506 + 1 \cdot 0.82494) = 1.4159$$

Ερώτημα β

Μέσω του παρακάτω κώδικα προσομοιώνουμε την διαδικασία. Δημιουργώ μια συνάρτηση με όρισμα τον αριθμό των σημείων που θα χρησιμοποιηθούν. Στη συνέχεια αφού οριστούν οι εκ των προτέρων πιθανότητες, παράγονται δείγματα μέσω γεννήτριας τυχαίων αριθμών από κανονική κατανομή με τις παραμέτρους που έχουμε. Αξίζει να σημειωθεί ότι ο αριθμός δειγμάτων της κάθε κλάσης θα αποτελεί ένα ποσοστό των συνολικών δειγμάτων ανάλογα με τις a priori πιθανότητες. Στη συνέχεια αρχικοποιώ σε μηδέν του αριθμούς των σημείων που ταξινομούνται ορθά και λανθασμένα σε κάθε κλάση και με χρήση των ορίων απόφασης(decision boundaries) που υπολογίστηκαν παραπάνω γίνεται ταξονόμηση των δειγμάτων και όταν ολοκλήρωθεί αυτό διαιρούμε τα αποτελέσματα με το αριθμό δειγμάτων, ώστε να προκύψουν πιθανότητες. Τέλος, υπολογίζουμε το ολικό ρίσκο και τυπώνουμε τα αποτελέσματα. Αφού, είναι έτοιμη η συνάρτηση αυτό που απομένει είναι να την καλέσουμε με τον επιθυμητό αριθμό συνολικών δειγμάτων.

```
import numpy as np
rng = np.random.default_rng(seed=123)



def gauss_decision(number_of_samples):
    p_omega1, p_omega2 = 1/3, 2/3
    w1_samples_size = int(p_omega1 * number_of_samples)
    w2_samples_size = int(p_omega2 * number_of_samples)
    l_11, l_12, l_21, l_22 = 1, 2, 3, 1
    # Generating samples from normal distribution
    w1 = rng.normal(2, np.sqrt(0.5), w1_samples_size)
    w2 = rng.normal(1.5, np.sqrt(0.2), w2_samples_size)
    w1_as_w2, w1_as_w1, w2_as_w1, w2_as_w2 = 0, 0, 0, 0
    # Using Decision Boundaries to classify data
    for p in w1:
        if 0.402986090944180 < p < 1.930347242389154:
            w1_as_w2 += 1
        else:
            w1_as_w1 += 1
    for p in w2:
        if 0.402986090944180 < p < 1.930347242389154:
            w2_as_w2 += 1
        else:
            w2_as_w1 += 1
```

```python
    # Calculating Probabilities
    p_w1_as_w1 = w1_as_w1 / w1_samples_size
    p_w1_as_w2 = w1_as_w2 / w1_samples_size
    p_w2_as_w1 = w2_as_w1 / w2_samples_size
    p_w2_as_w2 = w2_as_w2 / w2_samples_size
    # Total Cost
    cost = p_omega1 * ((l_11 * p_w1_as_w1) + (l_21 * p_w1_as_w2)) + \
           p_omega2 * ((l_12 * p_w2_as_w1) + (l_22 * p_w2_as_w2))
    # Print results
    print("Percentage of w1 classified as w1:", p_w1_as_w1)
    print("Percentage of w1 classified as w2:", p_w1_as_w2)
    print("Percentage of w2 classified as w1:", p_w2_as_w1)
    print("Percentage of w2 classified as w2", p_w2_as_w2)
    print("The cost is:", cost)


gauss_decision(300000)
```

Τα αποτελέσματα είναι τα εξής για συνολικό αριθμό δειγμάτων 300000 (όπου το $\frac{1}{3}$ είναι της πρώτης κλάσης και τα $\frac{2}{3}$ της 2ης). Όπως, παρατηρούμε το αποτέλεσμα είναι πολύ κοντα στα θεωρητικά με ακρίβεια 4ου δεκαδικού. Προφανώς εδώ οι υπολογισμοί γίνονται με στρογγυλοποίηση σε πολύ περισσοτερα δεκαδικά και οχι στα 4-5 όπου έγιναν θεωρητικά. Από το αρχείο HW2_Exercise3_4_visualizations.py παίρνουμε αντίστοιχα οπτικοποιημένα τα αποτελέσματα. Όπως βλέπουμε η καμπύλη απόφασης στο δεξί μέρος δεν ειναι απόλυτα συμμετρική και αυτό ο ωφείλεται στα $\lambda_{12} = 2, \lambda_{21} = 3$ δηλαδή το γεγονός ότι το ρίσκο στη μία περίπτωση λάθους είναι μεγαλύτερο απ ότι στην άλλη αλλά και στις διαφορετικές εκ των προτέρων πιθανότητες $P(\omega_1) = \frac{1}{3}$ και $P(\omega_2) = \frac{2}{3}$.

```
Percentage of w1 classified as w1: 0.5509
Percentage of w1 classified as w2: 0.4491
Percentage of w2 classified as w1: 0.17479
Percentage of w2 classified as w2 0.82521
The cost is: 1.4159266666666666
```