HW #3

① [Probabilistic generative models]

Generate $N_1 = 200$ datapoints from $\mathcal{N}(\vec{x} \mid \vec{\mu}_1, \Sigma)$ and $N_2 = 400$ datapoints from $\mathcal{N}(\vec{x} \mid \vec{\mu}_2, \Sigma)$, where

$\vec{\mu}_1 = (2, 2)$ and $\vec{\mu}_2 = (-2, -2)$.

$$\Sigma = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix} \Leftarrow \sigma = 2$$

Assume that the class-conditional densities are given by

$$\begin{cases} p(\vec{x} \mid C_1) = \mathcal{N}(\vec{x} \mid \vec{\mu}_1, \Sigma), \\ p(\vec{x} \mid C_2) = \mathcal{N}(\vec{x} \mid \vec{\mu}_2, \Sigma). \end{cases}$$

(a) Using exact values of $\vec{\mu}_1, \vec{\mu}_2, \Sigma, N_1, N_2$, find the exact decision boundary (DB) at which $p(C_1 \mid \vec{x}) = p(C_2 \mid \vec{x}) = 0.5$.
Plot the class-conditional densities and the exact DB as a heatmap or a contour map.

(b) Estimate $p(C_1 \mid \mathcal{D})$, $p(C_2 \mid \mathcal{D})$, $\vec{\mu}_1, \vec{\mu}_2, \Sigma$ by ML and find the DB using these estimated values. Add the DB to the plot in (a).

(2.) [Logistic regression]

Use the dataset from problem 1 and find the DB by logistic regression. Add it to the plot created for problem 1 and comment on whether the DB found by logistic regression corresponds to the other two (exact DB and the ML DB).

(3.) Bishop 4.19