DS 312 Project 4 Questions 1 & 2:

Question 1: Compute the maximum likelihood estimates for the mean μ_c and covariance matrix Σ_c for class c of a Bayesian classifier with Gaussian class-conditionals. You are given a set of N_c objects belonging to class c, represented as $x_1, x_2, \ldots, x_{N_c}$. Derive the formulas for these estimates.

In Bayesian Classifier with Gaussian class-conditionals, the data from each class c is assumed to follow a multivariate Gaussian distribution, given by

$$p(x_{N_c} \mid c) = \frac{1}{(2\pi)^{d/2} |\Sigma_c|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1}(x-\mu_c)\right)$$

where

- $x \in \mathbb{R}^d$ is the feature vector (data points) with d dimensions
- $\mu_c \in \mathbb{R}^d$ is the mean of class c
- $\Sigma_c \in \mathbb{R}^{d \times d}$ is the covariance matrix of class c.

Liklihood function for $x_1, x_2, \ldots, x_n$ is

$$P(x_{N_c} \mid T_{N_c} = c, X, T) = L(\mu_c, \Sigma_c) = \prod_{i=1}^{N_c} p(x_i \mid c)$$

$$\log L(\mu_c, \Sigma_c) = \sum_{i=1}^{N_c} \log\left(\frac{1}{(2\pi)^{d/2}|\Sigma_c|^{1/2}} \exp\left(-\frac{1}{2}(x_i-\mu_c)^T \Sigma_c^{-1}(x_i-\mu_c)\right)\right)$$

Simplifying,

$$\log L(\mu_c, \Sigma_c) = \frac{N_c}{2}\log|\Sigma_c| - \frac{N_c d}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N_c}(x_i-\mu_c)^T \Sigma_c^{-1}(x_i-\mu_c)$$

To estimate μ_c, differentiate log-likelihood & set it to 0.

$$\frac{\partial \log L(\mu_c, \Sigma_c)}{\partial \mu_c} = -0 \quad -0 + \Sigma_c^{-1}\sum_{i=1}^{N_c}(x_i-\mu_c) = 0$$

$$\sum_{i=1}^{N_c}(x_i-\mu_c) = 0$$

$$\boxed{\mu_c = \frac{1}{N_c}\sum_{i=1}^{N_c} x_i}$$

To estimate covariance matrix Σ_c, differentiate log-likelihood & set equal to 0.

$$\frac{\partial \log L(\mu_c, \Sigma_c)}{\partial \Sigma_c} = -\frac{N_c}{2}\Sigma_c^{-1} + \frac{1}{2}\Sigma_c^{-1}\left(\sum_{i=1}^{N_c}(x_i-\mu_c)(x_i-\mu_c)^T\right)\Sigma_c^{-1} = 0$$

$$\boxed{\Sigma_c = \frac{1}{N_c}\sum_{i=1}^{N_c}(x_i-\mu_c)(x_i-\mu_c)^T}$$

Question 2: Suppose the logistic regression model is defined as $y = w^T X$, where $y \in \{0, 1\}$ and w is a vector of coefficients and X represents the independent variables. Derive the logistic regression equation for $P(y=1 \mid X) = \frac{1}{1+e^{-w^T X}}$. Show all necessary steps in the derivation process.

$P(y=1 \mid X) = \frac{1}{1+e^{-w^T X}}$ is derived from the log-odd (logit) of $P(y=1 \mid X)$, which is given by:

$$\log\left(\frac{P(y=1 \mid X)}{P(y=0 \mid X)}\right) = \log\left(e^{w^T X}\right)$$

from definition of log-odds:

$$\frac{P(y=1 \mid X)}{P(y=0 \mid X)} = e^{w^T X}$$

Since $P(y=1 \mid X) + P(y=0 \mid X) = 1$, $\quad P(y=0 \mid X) = 1 - P(y=1 \mid X)$

$$\frac{P(y=1 \mid X)}{1-P(y=1 \mid X)} = e^{w^T X}$$

$$\frac{P(y=1 \mid X)(1-P(y=1 \mid X))}{1-P(y=1 \mid X)} = e^{w^T X}(1-P(y=1 \mid X))$$

$$P(y=1 \mid X) = e^{w^T X} - e^{w^T X} P(y=1 \mid X)$$

$$P(y=1 \mid X) + e^{w^T X} P(y=1 \mid X) = e^{w^T X}$$

$$P(y=1 \mid X)(1+e^{w^T X}) = e^{w^T X}$$

$$P(y=1 \mid X) = \frac{e^{w^T X}}{1+e^{w^T X}}$$

$$\boxed{\therefore P(y=1 \mid X) = \frac{1}{1+e^{-w^T X}}}$$