Problem 1

The Multidemensional Normal distribution is defined:

$$\mathcal{N}(\underline{x} \mid \underline{\mu}, \Sigma) \triangleq \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \cdot \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right]$$

where $\underline{\mu} = \mathbb{E}[\underline{x}] \in \mathbb{R}^D$ is the mean vector

$\Sigma = \text{Cov}[\underline{x}]$ is the $D \times D$ covariance matrix.

Σ is symetric, positive semi definite matrix.

$NLL(\underline{\theta}) = NLL(\underline{\mu}, \Sigma)$ (Negative log likelihood).

Let $\{\underline{X}_1, \dots, \underline{X}_N\}$: $\underline{X}_i \sim \mathcal{N}_N(\underline{x}; \underline{\mu}, \Sigma)$

By the independence of the r.v $\underline{X}_i$, the joint density of the data $\underline{X}_i, i=1,\dots,N$ is: $\prod_{i=1}^{N} \mathcal{N}_i(\underline{x}_i; \underline{\mu}, \Sigma)$

$$NLL(\underline{\mu}, \Sigma) = -\log \prod_{i=1}^{N} \mathcal{N}_i(\underline{x}_i \mid \underline{\mu}, \Sigma) = -\sum_{i=1}^{N} \log \mathcal{N}_i(\underline{x}_i; \underline{\mu}, \Sigma) =$$

$$= -\sum_{i=1}^{N} \log\left\{\frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x}_i-\underline{\mu})^T \Sigma^{-1} (\underline{x}_i-\underline{\mu})\right]\right\} =$$

$$= -\sum_{i=1}^{N}\left(-\frac{D}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\underline{x}_i-\underline{\mu})^T \Sigma^{-1} (\underline{x}_i-\underline{\mu})\right) =$$

$$= \frac{ND}{2}\log(2\pi) + \frac{N}{2}\log|\Sigma| + \sum_{i=1}^{N} \frac{1}{2}(\underline{x}_i-\underline{\mu})^T \Sigma^{-1} (\underline{x}_i-\underline{\mu})$$

Minimizing $NLL(\underline{\mu}, \Sigma)$ with respect to $\underline{\mu}$ will give us the MLE for the distribution's mean.

We know that: $\frac{\partial(\underline{w}^T A \underline{w})}{\partial \underline{w}} = 2A\underline{w}$ if $\underline{w}$ does not depend on A AND A is symetric

$$\frac{\partial NLL(\underline{\mu}, \Sigma)}{\partial \underline{\mu}} = 0 \Rightarrow \frac{\partial}{\partial \mu} \sum_{i=1}^{N} \frac{1}{2} (x_i - \underline{\mu})^T \Sigma^{-1} (\underline{x}_i - \mu) = 0 \Rightarrow$$

$$\sum_{i=1}^{N} \frac{\partial}{\partial \mu} \left[(\underline{x}_i - \underline{\mu})^T \Sigma^{-1} (\underline{x}_i - \underline{\mu}) \right] = 0 \overset{\circledast}{\Rightarrow}$$

$$\sum_{i=1}^{N} 2\Sigma^{-1}(\underline{x}_i - \underline{\mu}) = 0 \Rightarrow \sum_{i=1}^{N} (\underline{x}_i - \underline{\mu}) = 0 \Rightarrow$$

$$\sum_{i=1}^{N} \underline{x}_i - \sum_{i=1}^{N} \underline{\mu} = 0 \Rightarrow N \underline{\mu} = \sum_{i=1}^{N} \underline{x}_i \Rightarrow$$

$$\boxed{\mu_{MLE} = \frac{1}{N} \sum_{i=1}^{N} \underline{x}_i = \bar{\mu}}$$

The MLE of $\underline{\mu}$ is just the empirical mean.

Problem 2

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \cdots, n$$

- Binomial Theorem: $(a+\beta)^m = \sum_{i=0}^{m} \binom{m}{i} a^i \beta^{m-i}$

Calculate the mean:

$$\mathbb{E}[X] = \sum_x x\, p(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} =$$

$$= 0 + \sum_{x=1}^{n} \cancel{x} \frac{n!}{\cancel{x}!\,(n-x)!} p^x (1-p)^{n-x} =$$

$$= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!\,(n-x)!} p \cdot p^{x-1} (1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

Let: $\left.\begin{array}{l}(n-1) = a \\ (x-1) = \beta\end{array}\right\} \Rightarrow n - x = a - \beta$

Then: $x = 1 \Rightarrow \beta = 0$

$x = n \Rightarrow \beta = n - 1 = a$

$$\mathbb{E}[X] = np \sum_{\beta=0}^{a} \frac{a!}{\beta!\,(a-\beta)!} p^\beta (1-p)^{a-\beta} = np \sum_{\beta=0}^{a} \binom{a}{\beta} p^\beta (1-p)^{a-\beta}$$

$\xrightarrow{\text{Binomial Theorem}}$ $\mathbb{E}[X] = np(\cancel{p} + 1 - \cancel{p})^a \Rightarrow$

$$\boxed{\mathbb{E}[X] = np}$$

Calculate the variance

$$\mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 =$$

$$= \mathbb{E}[X^2] - (np)^2$$

We will calculate $\mathbb{E}[X^2]$.

$$\mathbb{E}[X^2] = \sum_{x=0}^{n} x^2 p(x) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} x \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} =$$

$$= np \sum_{x=1}^{n} x \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x}$$

Let: $\left.\begin{array}{l} n-1 = a \\ x-1 = \beta \end{array}\right\} \Rightarrow n-x = a-\beta$

Then: $x = 1 \Rightarrow \beta = 0$

$x = n \Rightarrow \beta = n-1 = a$

$$\mathbb{E}[X^2] = np \sum_{\beta=0}^{a} \frac{a!}{\beta!(a-\beta)!} p^\beta (1-p)^{a-\beta} \cdot (\beta+1) =$$

$$= np \sum_{\beta=0}^{a} (\beta+1) \binom{a}{\beta} p^\beta (1-p)^{a-\beta} =$$

$$= np \left[\sum_{\beta=0}^{a} \beta \binom{a}{\beta} p^\beta (1-p)^{a-\beta} + 1 \right] =$$

$$= np\left[\mathbb{E}[B] + 1\right] \xlongequal{B \sim \mathrm{Bin}(\beta|a,p)} np(ap+1) = np\left[(n-1)p+1\right]$$

It is:

$$\mathrm{Var}[x] = \mathbb{E}[x^2] - (np)^2 =$$

$$= np(n-1)p + np - (np)^2$$

$$= np(np-p) + np - (np)^2 =$$

$$= (np)^2 - np^2 + np - (np)^2 =$$

$$= np - np^2 = np(1-p)$$

$$\Rightarrow \mathrm{Var}[x] = np(1-p)$$

Problem 3

3.1. $\underline{X} = \{x_1, \dots, x_N\}$

The likelihood, that is the probability of the observed data data given μ, as a function of μ, is given by:

$$p(\underline{X}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$$

The (conjugate) prior distribution is:

$$p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2) \triangleq \frac{1}{(2\pi\sigma_0^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

The posterior distribution is given by:

$$p(\mu|\underline{X}) \propto p(\underline{X}|\mu) \cdot p(\mu) =$$

$$= \left[\frac{1}{(2\pi\sigma^2)^{N/2} \cdot (2\pi\sigma_0^2)^{1/2}}\right] \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n^2 - 2x_n\mu + \mu^2) - \frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right\} =$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{n=1}^{N}x_n^2 - 2\mu\sum_{n=1}^{N}x_n + N\mu^2\right) - \frac{1}{2\sigma_0^2}(\mu^2+\mu_0^2-2\mu\mu_0)\right\}$$

$$= \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{n=1}^{N}x_n^2 - 2\mu N\bar{x} + N\mu^2\right) - \frac{1}{2\sigma_0^2}(\mu^2+\mu_0^2-2\mu\mu_0)\right\}$$

$$= \exp\left\{-\frac{\sum x_n^2}{2\sigma^2} + \frac{\mu N\bar{x}}{\sigma^2} - \frac{N\mu^2}{2\sigma^2} - \frac{\mu^2}{2\sigma_0^2} - \frac{\mu_0^2}{2\sigma_0^2} + \frac{\mu\mu_0}{\sigma_0}\right\}$$

$$= \exp\left\{ -\frac{\mu^2}{2}\left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right) + \mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{N\bar{x}}{\sigma^2}\right) - \left(\frac{\mu_0^2}{2\sigma_0^2} + \frac{\sum x_n^2}{2\sigma^2}\right)\right\}$$

$$= \exp\left\{ -\frac{\mu_0^2}{2\sigma_0^2} - \frac{\sum_n x_n^2}{2\sigma^2}\right\} \exp\left\{ -\frac{\mu^2}{2}\left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right) + \mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{N\bar{x}}{\sigma^2}\right)\right\}$$

$$\propto \exp\left\{ -\frac{\mu^2}{2}\left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}\right) + \mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{N\bar{x}}{\sigma^2}\right)\right\}$$

$$= \exp\left\{ \mu^2\left(-\frac{1}{2\sigma_0^2} - \frac{N}{2\sigma^2}\right) + 2\mu\left(\frac{\mu_0}{2\sigma_0^2} + \frac{N\bar{x}}{2\sigma^2}\right)\right\}$$

We have: $\mu^2 \underbrace{\left(-\frac{1}{2\sigma_0^2} - \frac{N}{2\sigma^2}\right)}_{a} + 2\mu \underbrace{\left(\frac{\mu_0}{2\sigma_0^2} + \frac{N\bar{x}}{2\sigma^2}\right)}_{\beta} =$

$$= a\mu^2 + 2\mu\beta = a\left(\mu^2 + 2\frac{\beta}{a}\mu\right) = a\left(\mu^2 + 2\frac{\beta}{a}\mu + \left(\frac{\beta}{a}\right)^2 - \left(\frac{\beta}{a}\right)^2\right)$$

$$= a\left(\left(\mu + \frac{\beta}{a}\right)^2 - \frac{\beta^2}{a^2}\right)$$

(Completing the square trick).

Using the above, we now have:

$$p(\mu \mid X) \propto \exp\{a\mu^2 + 2\mu\beta\} = \exp\left\{a\left(\mu + \frac{\beta}{a}\right)^2 - \frac{\beta^2}{a}\right\}$$

$$\propto \exp\left\{a\left(\mu + \frac{\beta}{a}\right)^2\right\}$$

We can solve for the posterior parameters: μ_N and σ_N^2:

$$a = -\frac{1}{2\sigma_0^2} - \frac{N}{2\sigma^2} = \frac{-(\sigma^2 + \sigma_0^2 N)}{2\sigma_0^2\sigma^2} \Rightarrow$$

$$-\frac{1}{2\sigma_N^2} = -\frac{\sigma^2 + \sigma_0^2 N}{2\sigma_0^2\sigma^2} \rightarrow \boxed{\sigma_N^2 = \frac{\sigma_0^2\sigma^2}{N\sigma_0^2 + \sigma^2}}$$

$$\mu_N = -\frac{B}{a} = \left(\frac{\mu_0}{2\sigma_0^2} + \frac{N\bar{x}}{2\sigma^2}\right) \Big/ \left(\frac{1}{2\sigma_0^2} + \frac{N}{2\sigma^2}\right) =$$

$$= \frac{\dfrac{\cancel{2}\sigma^2\mu_0 + 2\sigma_0^2 N\bar{x}}{\cancel{2\sigma_0^2\sigma^2}}}{\dfrac{2\sigma^2 + 2\sigma_0^2 N}{\cancel{2\sigma_0^2\sigma^2}}} = \frac{N\sigma_0^2\bar{x} + \sigma^2\mu_0}{\sigma^2 + \sigma_0^2 N} \Leftrightarrow$$

$$\boxed{\mu_N = \frac{N\sigma_0^2\bar{x} + \sigma^2\mu_0}{\sigma^2 + \sigma_0^2 N}}$$