Question 1

Likelihood: $P(X|\mu) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2)$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x_n-\mu)^2}{\sigma^2}\right)$$

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

$$= C_1 \exp\left(-\frac{1}{2}\frac{\sum_{n=1}^{N}x_n^2+\sum_{n=1}^{N}\mu^2-2\mu\sum_{n=1}^{N}x_n}{\sigma^2}\right)$$

$$= C_1 \exp\left(-\frac{1}{2}\frac{\sum_{n=1}^{N}x_n^2+N\mu^2-2N\mu\mu_{ML}}{\sigma^2}\right)$$

μ_{ML}: maximum likelihood solution for μ given N data points. $\mu_{ML} = \frac{\sum_{n=1}^{N} x_n}{N}$

Prior: $P(\mu|\mu_0) = \mathcal{N}(\mu|\mu_0, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right) = C_2 \exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)$

Posterior: $P(\mu|X) \propto P(X|\mu)P(\mu|\mu_0)$

$$= C_1C_2 \exp\left[-\frac{1}{2}\left(\frac{\sum_{n=1}^{N}x_n^2+N\mu^2-2N\mu\mu_{ML}}{\sigma^2}+\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right]$$

considering only what is inside the parenthese:

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

we don't need to consider the terms without μ since they can be considered as ~~parts of norma~~ constants. thus the equations becomes:

$$-\frac{1}{2}\cdot\frac{1}{\sigma^2\sigma_0^2}\left[\sigma_0^2N\mu^2-2N\sigma_0^2\mu_{ML}\mu+\sigma^2\mu^2-2\mu_0\sigma^2\mu\right]$$

$$= -\frac{1}{2}\left[\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2-2\left(\frac{N\mu_{ML}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)\mu\right]$$

completing the square: when we have ax^2+2bx, we want to obtain $(x-c)^2$ by adding some constant.

here we have $a = \frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}$, $b = -\left(\frac{N\mu_{ML}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)$, thus we can write

$$-\frac{1}{2}(a\mu^2+2b\mu) = -\frac{a}{2}\left(\mu^2+\frac{2b}{a}\mu\right) = -\frac{a}{2}\left(\mu^2+\frac{2b}{a}\mu+\left(\frac{b}{a}\right)^2-\left(\frac{b}{a}\right)^2\right)$$

$= -\frac{a}{2}\left(\mu+\frac{b}{a}\right)^2+\frac{b^2}{2a}$ the term $\exp\left(\frac{b^2}{2a}\right)$ is also a constant, we can discard it.

$\therefore$ The term inside the paranthese of the exponential is

$$-\frac{1}{2}\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\left[\mu-\left(\frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0+\frac{\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}\right)\right]^2$$

With $\mu_N = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}$, $\sigma_N^2 = \left(\frac{1}{\sigma_0^2}+\frac{N}{\sigma^2}\right)^{-1}$, we can see that

the posterior $p(\mu|x) \propto C\exp\left(-\frac{1}{2}\frac{(\mu-\mu_N)^2}{\sigma_N^2}\right)$. qed.

Question 2

Likelihood: $p(X|\vec{\mu},\Sigma) = \prod_{n=1}^{N}\mathcal{N}(\vec{x}_n|\vec{\mu},\Sigma)$ assuming l components of multivariate gaussing

$$= \prod_{n=1}^{N}\frac{1}{(2\pi)^{\frac{l}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\vec{x}_n-\vec{\mu})^T\Sigma^{-1}(\vec{x}_n-\vec{\mu})\right]$$

Prior: $p(\vec{\mu}|\vec{\mu}_0,\Sigma_0) = \frac{1}{(2\pi)^{\frac{l}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\vec{\mu}-\vec{\mu}_0)^T\Sigma_0^{-1}(\vec{\mu}-\vec{\mu}_0)\right]$

Posterior: $p(\vec{\mu}|X) \propto C_1C_2\exp\left[\sum_{n=1}^{N}\left(-\frac{1}{2}(\vec{x}_n-\vec{\mu})^T\Sigma^{-1}(\vec{x}_n-\vec{\mu})\right)-\frac{1}{2}(\vec{\mu}-\vec{\mu}_0)^T\Sigma_0^{-1}(\vec{\mu}-\vec{\mu}_0)\right]$

Consider only what's inside the parenthese of the exponential:

$$-\frac{1}{2}\left[\sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})^T\Sigma^{-1}(\vec{x}_n-\vec{\mu})+(\vec{\mu}-\vec{\mu}_0)^T\Sigma_0^{-1}(\vec{\mu}-\vec{\mu}_0)\right]$$

$$= -\frac{1}{2}\left[\sum_{n=1}^{N}\vec{x}_n^T\Sigma^{-1}\vec{x}_n-\sum_{n=1}^{N}\vec{x}_n^T\Sigma^{-1}\vec{\mu}-\sum_{n=1}^{N}\vec{\mu}^T\Sigma^{-1}\vec{x}_n+\sum_{n=1}^{N}\vec{\mu}^T\Sigma^{-1}\vec{\mu}+\vec{\mu}^T\Sigma_0^{-1}\vec{\mu}-\vec{\mu}^T\Sigma_0^{-1}\vec{\mu}_0-\vec{\mu}_0^T\Sigma_0^{-1}\vec{\mu}+\vec{\mu}_0^T\Sigma_0^{-1}\vec{\mu}_0\right]$$

As we did before, we can discard terms without $\vec{\mu}$, thus:

$$-\frac{1}{2}\left[-N\vec{\mu}_{ML}^T\Sigma^{-1}\vec{\mu}-N\vec{\mu}^T\Sigma^{-1}\vec{\mu}_{ML}+N\vec{\mu}^T\Sigma^{-1}\vec{\mu}+\vec{\mu}^T\Sigma_0^{-1}\vec{\mu}-\vec{\mu}^T\Sigma_0^{-1}\vec{\mu}_0-\vec{\mu}_0^T\Sigma_0^{-1}\vec{\mu}\right]$$

where $\vec{\mu}_{ML}$ is the maximum likelihood for $\vec{\mu}$ given N data points. $\vec{\mu}_{ML} = \frac{\sum_{n=1}^{N}\vec{x}_n}{N}$

$$= -\frac{1}{2}\left[\vec{\mu}^T(N\Sigma^{-1}+\Sigma_0^{-1})\vec{\mu}-(N\vec{\mu}_{ML}^T\Sigma^{-1}+\vec{\mu}_0^T\Sigma_0^{-1})\vec{\mu}-\vec{\mu}^T(N\Sigma^{-1}\vec{\mu}_{ML}+\Sigma_0^{-1}\vec{\mu}_0)\right]$$

Let $A = N\Sigma^{-1}+\Sigma_0^{-1}$, $\vec{b} = N\Sigma^{-1}\vec{\mu}_{ML}+\Sigma_0^{-1}\vec{\mu}_0$. thus $\vec{b}^T = N\vec{\mu}_{ML}^T\Sigma^{-1}+\vec{\mu}_0^T\Sigma_0^{-1}$, thus

$$= -\frac{1}{2}\left[\vec{\mu}^TA\vec{\mu}-\vec{b}^T\vec{\mu}-\vec{\mu}^T\vec{b}\right]$$

$\because$ A is the weighted sum of ~~the~~ two symmetric, full-rank covariance matrices.

$\therefore$ A is symmetric and invertible $\therefore A^{-1}A = AA^{-1} = I$

$$\therefore -\frac{1}{2}(\vec{\mu}^TA\vec{\mu}-\vec{b}^T\vec{\mu}-\vec{\mu}^T\vec{b}) = -\frac{1}{2}(\vec{\mu}^TA\vec{\mu}-\vec{b}^TA^{-1}A\vec{\mu}-\vec{\mu}^TAA^{-1}\vec{b}+\vec{b}^TA^{-1}AA^{-1}\vec{b})+\frac{1}{2}\vec{b}^TA^{-1}\vec{b}$$

$\frac{1}{2}\vec{b}^TA^{-1}\vec{b}$ is a constant with regard to $\vec{\mu}$, we can discard this term

Let $\Sigma_N = A^{-1}$, $\vec{\mu}_N = A^{-1}b$, we got

$$-\frac{1}{2}(\vec{\mu}^T\Sigma_N^{-1}\vec{\mu}-\vec{\mu}_N^T\Sigma_N^{-1}\vec{\mu}-\vec{\mu}^T\Sigma_N^{-1}\vec{\mu}_N+\vec{\mu}_N^T\Sigma_N^{-1}\vec{\mu}_N) = -\frac{1}{2}(\vec{\mu}-\vec{\mu}_N)^T\Sigma_N^{-1}(\vec{\mu}-\vec{\mu}_N)$$

$\therefore p(\vec{\mu}|X) \propto \exp\left[-\frac{1}{2}(\vec{\mu}-\vec{\mu}_N)^T\Sigma_N^{-1}(\vec{\mu}-\vec{\mu}_N)\right]$.

$\vec{\mu}_N = A^{-1}\vec{b} = (N\Sigma^{-1}+\Sigma_0^{-1})^{-1}(N\Sigma^{-1}\vec{\mu}_{ML}+\Sigma_0^{-1}\vec{\mu}_0)$

$\Sigma_N = A^{-1} = (N\Sigma^{-1}+\Sigma_0^{-1})^{-1}$ Qed