Shashank Singh[1]

# 4 Advanced Theory

### Part A [15 points]

1. By properties of the logarithm,

$$\ell(\mu,\Sigma) = \sum_{i=1}^n \log \mathcal{N}(x_i;\mu,\Sigma) = -\sum_{i=1}^n \log\left((2\pi)^{n/2}\sqrt{\det\Sigma}\right) + \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu).$$

2.

$$\begin{aligned}
\frac{\partial}{\partial\mu}\ell(\mu,\Sigma) &= -\frac{\partial}{\partial\mu}\sum_{i=1}^n \log(2\pi)^{n/2}\sqrt{\det\Sigma} + \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu). \\
&= -\sum_{i=1}^n \frac{1}{2}\frac{\partial}{\partial\mu}\left(x_i^T\Sigma^{-1}x_i - \mu^T\Sigma^{-1}x_i - x_i^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right) \\
&= \sum_{i=1}^n \frac{\partial}{\partial\mu}\left(\mu\Sigma^{-1}x_i^T - \frac{\mu^T\Sigma^{-1}\mu}{2}\right) \\
&= \sum_{i=1}^n \Sigma^{-1}x_i^T - \Sigma^{-1}\mu = \left(\sum_{i=1}^n \Sigma^{-1}x_i^T\right) - n\Sigma^{-1}\mu.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\Sigma^{-1}}\ell(\mu,\Sigma) &= -\frac{\partial}{\partial\Sigma^{-1}}\sum_{i=1}^n \log\left((2\pi)^{n/2}\sqrt{\det\Sigma}\right) + \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu). \\
&= -\frac{n}{2}\frac{\partial}{\partial\Sigma^{-1}}\log(\det\Sigma) - \frac{\partial}{\partial\Sigma^{-1}}\sum_{i=1}^n \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) \\
&= \frac{n}{2}\frac{\partial}{\partial\Sigma^{-1}}\log(\det\Sigma^{-1}) - \frac{\partial}{\partial\Sigma^{-1}}\operatorname{tr}\left(\Sigma^{-1}\sum_{i=1}^n \frac{1}{2}(x_i-\mu)(x_i-\mu)^T\right) \\
&= \frac{n}{2}\Sigma - \sum_{i=1}^n \frac{1}{2}(x_i-\mu)(x_i-\mu)^T.
\end{aligned}$$

3. For any matrix $V \in \mathcal{S}_+^d$ define $g_V : \mathbb{R} \to \mathbb{R}$ by

$$g_V(t) = g(\Sigma + tV),$$

[1] sss1@andrew.cmu.edu

and note that it suffices to show that $g_V$ is concave (i.e., that $g_V''(t) \leq 0$) when $\Sigma + tV \succ 0$. Since $\Sigma \succ 0$, $\Sigma$ has a square root $\sqrt{\Sigma} \succ 0$.

$$\begin{aligned}
g_V(t) = g\left(\Sigma + tV\right) &= g\left(\sqrt{\Sigma}\left(I + t\sqrt{\Sigma}^{-1}V\sqrt{\Sigma}^{-1}\right)\sqrt{\Sigma}\right) \\
&= \log\left(\sqrt{\det\Sigma}\det\left(I + t\sqrt{\Sigma}^{-1}V\sqrt{\Sigma}^{-1}\right)\sqrt{\det\Sigma}\right) \\
&= \log\left(\det\Sigma\right) + \log\left(\det\left(I + t\sqrt{\Sigma}^{-1}V\sqrt{\Sigma}^{-1}\right)\right) \\
&= \log\left(\det\Sigma\right) + \sum_{i=1}^{n}\log\left(1 + t\lambda_i\right),
\end{aligned}$$

where $\lambda_1, \ldots, \lambda_n \geq 0$ are the eigenvalues of $\sqrt{\Sigma}^{-1}V\sqrt{\Sigma}^{-1}$. Differentiating twice then gives

$$g_V'(t) = \sum_{i=1}^{n}\frac{\lambda_i}{1 + t\lambda_i} \quad \text{and} \quad g_V''(t) = \sum_{i=1}^{n}\frac{-\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0.$$

Now define $f_i : \mathbb{R} \to \mathbb{R}$ by

$$f_i(t) = h_i((\mu_0, \Sigma_0) + t(\mu, \Sigma)).$$

From part 1.,

$$\begin{aligned}
\ell(\mu, \Sigma) &= -\sum_{i=1}^{n}\log\left((2\pi)^{n/2}\sqrt{\det\Sigma}\right) + \frac{1}{2}(x_i - \mu)^T\Sigma^{-1}(x_i - \mu) \\
&= -\sum_{i=1}^{n}\log\left((2\pi)^{n/2}\right) + \frac{1}{2}\log\left(\det\Sigma\right) + \frac{1}{2}(x_i - \mu)^T\Sigma^{-1}(x_i - \mu).
\end{aligned}$$

Since $g$ is concave and each $h_i$ is convex, $\ell(\mu, \Sigma)$ is generally neither convex nor concave. ■

4. Setting the derivatives in part 2 to 0 and solving for $\mu$ and $\Sigma$ gives

$$\mu_{mle} = \frac{1}{n}\sum_{i=1}^{n}x_i \quad \text{and} \quad \Sigma_{mle} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T.$$

Plugging these into the $\mathcal{N}(x; \mu, \Sigma)$ gives the MLE estimate of the Gaussian.

**Part B [10 points]** Didn't have time to finish this part.