1. Given

$$f(x) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)},$$

where $x, \mu \in \mathbb{R}^k$, Σ is a k-by-k positive definite matrix and $|\Sigma|$ is its determinant. Show that $\int_{\mathbb{R}^k} f(x)\,dx = 1$.

2. Let A, B be n-by-n matrices and x be a n-by-1 vector.
   (a) Show that $\frac{\partial}{\partial A}\mathrm{trace}(AB) = B^T$.
   (b) Show that $x^T A x = \mathrm{trace}(xx^T A)$.
   (b) Derive the maximum likelihood estimators for a multivariate Gaussian.

1.

$$Let\ z = \Sigma^{\frac{-1}{2}}(x-\mu)$$
$$dx = |\Sigma|^{\frac{-1}{2}}dz$$
$$\Rightarrow \int_{\mathbb{R}^k} f(x)dx = \int_{\mathbb{R}^k} \frac{1}{(2\pi)^{\frac{k}{2}}} e^{\frac{1}{2}z^T z}dz = 1_{(QED)}$$

2.(a)

$$trace(AB) = \sum_{i,j} A_{ij}B_{ji}$$
$$\Rightarrow \frac{\partial}{\partial A_{ij}} trace(AB) = B_{ji}$$
$$\Rightarrow \frac{\partial}{\partial A} trace(AB) = B^T_{(QED)}$$

2.(b)

$$x^T Ax = c,\ where\ c\ is\ a\ scalar$$
$$x^T Ax = trace(x^T Ax)$$
$$Due\ to\ the\ cyclic\ property\ of\ trace,\ x^T Ax = trace(xx^T A)_{(QED)}$$

2.(c)

$Let\ x_1, x_2, \ldots, x_n \in \mathbb{R}^k\ be\ i.i.d.\ samples\ on\ \mathcal{N}(\mu, \Sigma)$

$\mathcal{L}(\mu, \Sigma) = \sum_{i=1}^{n} \log f(x_i)$

$= -\frac{nk}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T\Sigma^{-1}(x_i - \mu)$

$\frac{\partial \mathcal{L}}{\partial \mu} = \Sigma^{-1}\sum_{i=1}^{n}(x_i - \mu) = 0$

$\Rightarrow \sum_{i=1}^{n}(x_i - \mu) = 0$

$\Rightarrow n\mu = \sum_{i=1}^{n} x_i$

$\Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$Therefore,\ \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T$

3.

I learned about generalizing linear model and the exponential family in class this week. It makes me think of how the exponential family form of other distribution will be like. So I have some search on the net and find its general form.