Viterbo Giuseppe Homework 2

Matricola: 2086516

February 8, 2024

1 EX. 4

Show that the characteristic function of a multivariate Gaussian distribution, $N(x|\mu, C)$, is equal to:

$$\phi(k) = exp(-i\mu^T \cdot k - \frac{1}{2}k^T Ck) \tag{1}$$

1.1 Proof: Completing the square

We need to start from the Fourier transform of the distribution:

$$\begin{aligned} \phi(k) &= \int d^n x \, exp(-ik \cdot x)p(x) \\ &= \int d^n x \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} exp\bigg(- i\vec{k} \cdot \vec{x} - \frac{1}{2}(\vec{x} - \vec{\mu})^T C^{-1}(\vec{x} - \vec{\mu})\bigg) \end{aligned} \tag{2}$$

and we need to apply the square completion formula to the exponential part with the following replacement:

$$\begin{aligned} A &= C^{-1} \\ \vec{z} &= (\vec{x} - \vec{\mu}) \\ \vec{b} &= i\vec{k} \\ c &= i\vec{k}^T \vec{\mu} \end{aligned}$$

and in this way the exponent in the integral becomes:

$$\begin{aligned} -i\vec{k} \cdot \vec{x} - \frac{1}{2}(\vec{x} - \vec{\mu})^T C^{-1}(\vec{x} - \vec{\mu}) &= -\frac{1}{2}\bigg((\vec{x} - \vec{\mu}) + C(i\vec{k})\bigg)^T C^{-1}\bigg((\vec{x} - \vec{\mu}) + C(i\vec{k})\bigg) + \frac{1}{2}(i\vec{k}^T)C(i\vec{k}) - i\vec{k} \cdot \vec{\mu}^T \\ &= -\frac{1}{2}\bigg(\vec{x} - \vec{\tilde{\mu}}\bigg)^T C^{-1}\bigg(\vec{x} - \vec{\tilde{\mu}}\bigg) - \frac{1}{2}\vec{k}^T C\vec{k} - i\vec{k} \cdot \vec{\mu}^T \end{aligned} \tag{3}$$

where in the last passage we have defined the $\vec{\tilde{\mu}} = \vec{\mu} - C(i\vec{k})$. Now we can go back to the integral and use the fact that the normalization constant for $N(x|\mu, C)$ and $N(x|\vec{\tilde{\mu}}, C)$ is the same:

$$
\begin{aligned}
\phi(k) &= \int d^n x \, exp(-ik\cdot x)p(x) \\
&= \int d^n x \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} exp\Bigg(-\frac{1}{2}\Big(\vec{x}-\vec{\mu}\Big)^T C^{-1}\Big(\vec{x}-\vec{\mu}\Big) - \frac{1}{2}\vec{k}^T C\vec{k} - i\vec{k}\cdot\vec{\mu}^T \Bigg) \\
&= \int d^n x \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} exp\Bigg(-\frac{1}{2}\Big(\vec{x}-\vec{\mu}\Big)^T C^{-1}\Big(\vec{x}-\vec{\mu}\Big)\Bigg) exp\Bigg(- \frac{1}{2}\vec{k}^T C\vec{k} - i\vec{k}\cdot\vec{\mu}^T \Bigg) \\
&= exp\Bigg(- \frac{1}{2}\vec{k}^T C\vec{k} - i\vec{k}\cdot\vec{\mu}^T \Bigg) \underbrace{\int d^n x \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} exp\Bigg(-\frac{1}{2}\Big(\vec{x}-\vec{\mu}\Big)^T C^{-1}\Big(\vec{x}-\vec{\mu}\Big)\Bigg)}_{=1} \\
&= exp\Bigg(- \frac{1}{2}\vec{k}^T C\vec{k} - i\vec{k}\cdot\vec{\mu}^T \Bigg)
\end{aligned} \tag{4}
$$

1.2 Proof: Rotation

Since the covariance matrix is symmetric it is possible to find a change of coordinates to diagonalize it, and in particular it is possible to find O such that $C^{-1} = O^T D^{-1} O$ with D^{-1} diagonal with eigenvalues d_i [1]. By defining $\vec{y} = O\vec{x}$, $\vec{\nu} = O\vec{\mu}$, $\vec{l} = O\vec{k}$, we can rewrite the exponential in the $\phi(k)$ in the following way:

$$
\begin{aligned}
-i\vec{k}\cdot\vec{x} - \frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu}) &= -i\vec{k}^T O^T O\vec{x} - \frac{1}{2}(\vec{x}-\vec{\mu})^T O^T D^{-1} O(\vec{x}-\vec{\mu}) \\
&= -i\vec{l}\cdot\vec{y} - \frac{1}{2}(\vec{y}-\vec{\nu})^T D^{-1}(\vec{y}-\vec{\nu}) \\
&= -\sum_i il_i y_i + \frac{1}{2}d_i(y_i-\nu_i)^2 \\
&= -\sum_i il_i y_i + \frac{d_i}{2}y_i^2 - d_i y_i \nu_i + \frac{d_i}{2}\nu_i^2 \\
&= -\sum_i \frac{d_i}{2}y_i^2 + y_i(il_i - d_i\nu_i) + \frac{d_i}{2}\nu_i^2 \\
&= -\sum_i \frac{d_i}{2}\left[y_i^2 + 2y_i\left(\frac{il_i - d_i\nu_i}{d_i}\right) + v_i^2 \right] \\
&= -\sum_i \frac{d_i}{2}\left[y_i^2 + 2y_i\left(\frac{il_i - d_i\nu_i}{d_i}\right) - \left(\frac{il_i - d_i\nu_i}{d_i}\right)^2 + \left(\frac{il_i - d_i\nu_i}{d_i}\right)^2 + v_i^2 \right] \\
&= -\sum_i \frac{d_i}{2}\left[\left(y_i - \frac{il_i - d_i\nu_i}{d_i}\right)^2 - \left(\frac{il_i - d_i\nu_i}{d_i}\right)^2 + v_i^2 \right] \\
&= \sum_i -\frac{d_i}{2}\left(y_i - \frac{il_i - d_i\nu_i}{d_i}\right)^2 + \left(\frac{(il_i - d_i\nu_i)^2}{2d_i} - \frac{d_i}{2}v_i^2\right)
\end{aligned} \tag{5}
$$

[1] In this notations the eigevalues of D^{-1} are d_i, while the eigevalues of D are d_i^{-1}

Now we can put this expression inside the integral to get the $\phi(k)$, taking advantage of the fact that $detO = 1$ so we have that $d^n x = d^n y$:

$$
\begin{aligned}
\phi(k) &= \int d^n x \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} exp\bigg(-i\vec{k}\cdot\vec{x} - \frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})\bigg) \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} \prod_i \int dy_i\, exp\Bigg(\sum_i -\frac{d_i}{2}\bigg(y_i - \frac{il_i - d_i\nu_i}{d_i}\bigg)^2 + \bigg(\frac{(il_i - d_i\nu_i)^2}{2d_i} - \frac{d_i}{2}\nu_i^2\bigg)\Bigg) \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} \prod_i \int dy_i\, exp\Bigg(-\frac{d_i}{2}\bigg(y_i - \frac{il_i - d_i\nu_i}{d_i}\bigg)^2\Bigg) exp\bigg(\frac{(il_i - d_i\nu_i)^2}{2d_i} - \frac{d_i}{2}\nu_i^2\bigg) \\
&= \Bigg[\prod_i exp\bigg(\frac{(il_i - d_i\nu_i)^2}{2d_i} - \frac{d_i}{2}\nu_i^2\bigg)\Bigg] \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} \prod_i \underbrace{\int dy_i\, exp\Bigg(-\frac{d_i}{2}\bigg(y_i - \frac{il_i - d_i\nu_i}{d_i}\bigg)^2\Bigg)}_{\int dx e^{-a(x-b)^2} = \sqrt{\frac{2\pi}{a}}} \\
&= exp\bigg(\sum_i \frac{(il_i - d_i\nu_i)^2}{2d_i} - \frac{d_i}{2}\nu_i^2\bigg) \frac{1}{(2\pi)^{n/2}\sqrt{det\, C}} \underbrace{\prod_i \sqrt{\frac{2\pi}{d_i}}}_{= \frac{(2\pi)^{n/2}}{\sqrt{det(C^{-1})}}} \\
&= exp\bigg(\sum_i \frac{-l_i^2 + d_i^2\nu_i^2 - 2l_i d_i \nu_i - d_i^2\nu_i^2}{2d_i}\bigg) \frac{1}{\sqrt{detC \cdot det(C^{-1})}} \\
&= exp\bigg(-\sum_i \frac{1}{2} l_i^2 d_i^{-1} + l_i\nu_i\bigg) \\
&= exp\bigg(-\frac{1}{2}\vec{l}^T D\vec{l} - \vec{l}\cdot\vec{\nu}\bigg) \\
&= exp\bigg(-\frac{1}{2}(O\vec{k})^T D(O\vec{k}) - (O\vec{k})\cdot(O\vec{\mu})\bigg) \\
&= exp\bigg(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\vec{\mu}^T\bigg)
\end{aligned} \tag{6}
$$

2 Ex. 5

We want to calculate the mean and standard deviation of the Multivariate Gaussian $N(\mu, \Sigma)$ using the characteristic function via differentiation as follow:

$$E\big[x_\alpha^{n_\alpha}...x_\beta^{n_\beta}\big] = \left[\frac{\partial^{n_\alpha...n_\beta}\phi(k)}{\partial(-ik_\alpha)^{n_\alpha}...\partial(-ik_\beta)^{n_\beta}}\right]_{k=0} \tag{7}$$

We can then calculate the mean component by component as follows:

$$\begin{aligned}
E(x_\alpha) &= \frac{\partial\phi(\vec{k})}{\partial(-i\vec{k}_\alpha)}\bigg|_{\vec{k}=0} \\
&= \frac{\partial}{\partial(-i\vec{k}_\alpha)}\bigg|_{\vec{k}=0} exp\Big(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\Big) \\
&= \frac{\partial}{\partial(-i\vec{k}_\alpha)}\bigg|_{\vec{k}=0} exp\Big(-\frac{1}{2}\sum_i\sum_j k_j C_{ij} k_i - \sum_t k_t\mu_t\Big) \\
&= \left[\frac{\partial}{\partial(-i\vec{k}_\alpha)}\Big(-\frac{1}{2}\sum_i\sum_j k_i C_{ij} k_j - \sum_t k_t\mu_t\Big)\right] exp\Big(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\Big)\bigg|_{\vec{k}=0} \\
&= \left[-\frac{1}{2}i\sum_j C_{\alpha j}k_j - \frac{1}{2}i\underbrace{\sum_i C_{i\alpha}k_i}_{C_{i\alpha}=C_{\alpha i}} + \mu_\alpha\right] exp\Big(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\Big)\bigg|_{\vec{k}=0} \\
&= \left[-i\sum_i C_{i\alpha}k_i + \mu_\alpha\right] exp\Big(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\Big)\bigg|_{\vec{k}=0} \\
&= \mu_\alpha
\end{aligned} \tag{8}$$

For the Covariance matrix $\tilde{C}_{\alpha\beta}$ we need to compute:

$$\begin{aligned}
\tilde{C}_{\alpha\beta} &= E\Big[(x_\alpha - E[x_\alpha])(x_\beta - E[x_\beta])\Big] \\
&= E\Big[x_\alpha x_\beta - x_\alpha E[x_\beta] - x_\beta E[x_\alpha] + E[x_\alpha]E[x_\beta]\Big] \\
&= E[x_\alpha x_\beta] - \mu_\alpha\mu_\beta - \mu_\alpha\mu_\beta + \mu_\alpha\mu_\beta \\
&= E[x_\alpha x_\beta] - \mu_\alpha\mu_\beta
\end{aligned} \tag{9}$$

We need to compute only $E[x_\alpha x_\beta]$ as follows:

$$
\begin{aligned}
E[x_\alpha x_\beta] &= \frac{\partial^2 \phi(\vec{k})}{\partial(-i\vec{k}_\alpha)\partial(-i\vec{k}_\beta)}\bigg|_{\vec{k}=0} \\
&= \frac{\partial}{\partial(-i\vec{k}_\beta)}\bigg|_{\vec{k}=0}\left[-i\sum_i C_{\alpha i}k_i + \mu_\alpha\right] exp\left(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\right) \\
&= \left[C_{\alpha\beta} + \mu_\alpha\left(-i\sum_i C_{i\beta}k_i + \mu_\beta\right)\right] exp\left(-\frac{1}{2}\vec{k}^T C\vec{k} - \vec{k}\cdot\vec{\mu}\right)\bigg|_{\vec{k}=0} \\
&= C_{\alpha\beta} + \mu_\alpha\mu_\beta
\end{aligned}
\tag{10}
$$

In the end we get that $\tilde{C}_{\alpha\beta} = C_{\alpha\beta}$.

3 Ex. 8

Let's start from some tensor notation:

- $\vec{d}$ has length N, and it's components are gonna be indicated as d_μ and d_ν;
- $\vec{A}$ are the M unknown amplitudes for the templates and it are gonna be indicated as A_i and A_j;
- T is the matrix that has for columns the M template vector, each of dimension N like $\vec{d}$, so it is gonna be indicated as $t_{k\alpha}$, with $k \in \{i, j\}$ and $\alpha \in \{\mu, \nu\}$;

Using also Einstein summation convention, we can rewrite the χ^2 as:

$$
\begin{aligned}
\chi^2 &= (\vec{d} - \vec{A}T)^t C^{-1}(\vec{d} - \vec{A}T) \\
&= \big(d_\mu - A_i t_{i\mu}\big) C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big) \\
&= \big(d_\mu - A_i t_{i\mu}\big) C^{-1}_{\mu\nu}\big(d_\nu - A_i t_{i\nu}\big) \\
&= \big(d_\mu - A_j t_{j\mu}\big) C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big)
\end{aligned}
\tag{11}
$$

In the last equation we have used the fact that, since the latin index are dummy index, we can rename it as we want. In order to maximise we need to take the first derivative of χ^2 with respect to A_l as follows:

$$
\begin{aligned}
\frac{\partial\chi^2}{\partial A_l} &= \frac{\partial}{\partial A_l}\left[\big(d_\mu - A_j t_{j\mu}\big) C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big)\right] \\
&= -t_{l\mu}\left(C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big)\right) - \left(\big(d_\mu - A_j t_{j\mu}\big) C^{-1}_{\mu\nu}\right) t_{l\nu} \\
&= C^{-1}_{\mu\nu}\Big[\underbrace{-t_{l\mu}\big(d_\nu - A_j t_{j\nu}\big) - t_{l\nu}\big(d_\mu - A_j t_{j\mu}\big)}_{= -2t_{l\mu}(d_\nu - A_j t_{j\nu})}\Big] \\
&= -2 t_{l\mu} C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big)
\end{aligned}
\tag{12}
$$

Since we want to find the maximum we need to verify that the second derivative of χ^2 is positive:

$$
\begin{aligned}
\frac{\partial^2\chi^2}{\partial A_l A_m} &= \frac{\partial}{\partial A_m}\Big[-2t_{l\mu}C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big)\Big] \\
&= 2t_{l\mu}C^{-1}_{\mu\nu}t_{m\nu} > 0
\end{aligned} \tag{13}
$$

In the last passage we used the fact that the inverse of the covariance matrix is positive definite. Since we now know that we are finding the maximum of the χ^2 function we can solve the following equation:

$$
\begin{aligned}
\frac{\partial\chi^2}{\partial A_l} &= 0 \\
-2t_{l\mu}C^{-1}_{\mu\nu}\big(d_\nu - A_j t_{j\nu}\big) &= 0 \\
-t_{l\mu}C^{-1}_{\mu\nu}d_\nu + t_{l\mu}C^{-1}_{\mu\nu}A_j t_{j\nu} &= 0 \\
t_{l\mu}C^{-1}_{\mu\nu}A_j t_{j\nu} &= t_{l\mu}C^{-1}_{\mu\nu}d_\nu \\
t_{l\mu}C^{-1}_{\mu\nu}t_{j\nu}A_j &= t_{l\mu}C^{-1}_{\mu\nu}d_\nu \\
\underbrace{\big[t_{k\mu}C^{-1}_{\mu\nu}t_{l\nu}\big]^{-1}t_{l\mu}C^{-1}_{\mu\nu}t_{j\nu}}_{=\delta_{kj}}A_j &= \big[t_{k\mu}C^{-1}_{\mu\nu}t_{j\nu}\big]^{-1}t_{l\mu}C^{-1}_{\mu\nu}d_\nu \\
A_k &= \big[t_{k\mu}C^{-1}_{\mu\nu}t_{l\nu}\big]^{-1}t_{l\mu}C^{-1}_{\mu\nu}d_\nu \\
\vec{A} &= [TC^{-1}T^t]^{-1}TC^{-1}\vec{d}
\end{aligned} \tag{14}
$$

So in the end we get a system of M equations for the M components of the amplitudes $\vec{A}$.

4 Ex. 9

We can solve the MAP estimate of parameter ω and b by using the following substitutions in the solution of the previous exercise:

$$
C = \sigma_d^2 \quad \vec{A} = \begin{bmatrix} w \\ b \end{bmatrix} \quad T = \begin{bmatrix} x_1 & ... & x_N \\ 1 & ... & 1 \end{bmatrix} \quad \vec{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}
$$

In this way the best estimate for the linear regression parameter can be computed by solving the final expression of equation 14. We need only to compute $[TC^{-1}T^t]^{-1}$ as follows:

$$
\begin{aligned}
[TC^{-1}T^t]^{-1} &= \left[\frac{1}{\sigma_d^2}\begin{bmatrix} x_1 & ... & x_N \\ 1 & ... & 1 \end{bmatrix}\begin{bmatrix} x1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}\right]^{-1} \\
&= \left[\frac{N}{\sigma_d^2}\begin{bmatrix} \bar{x^2} & \bar{x} \\ \bar{x} & 1 \end{bmatrix}\right]^{-1} \\
&= \frac{\sigma_d^2}{(N\sum_i x_i^2 - (\sum_i x_i)^2)}\begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & (\bar{x})^2 \end{bmatrix} \\
&= \frac{\sigma_d^2}{\Delta}\begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & (\bar{x})^2 \end{bmatrix}
\end{aligned}
\tag{15}
$$

Where in the last line we have simply introduce $\Delta = N\sum_i x_i^2 - (\sum_i x_i)^2$. Now we can obtain the estimate for $\vec{A}$:

$$
\begin{aligned}
\vec{A} &= [TC^{-1}T^t]^{-1}TC^{-1}\vec{d} \\
&= \frac{\sigma_d^2}{\sigma_d^2\Delta}\begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & (\bar{x})^2 \end{bmatrix}\begin{bmatrix} x_1 & ... & x_N \\ 1 & ... & 1 \end{bmatrix}\begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix} \\
\begin{bmatrix} w \\ b \end{bmatrix} &= \frac{1}{\Delta}\begin{bmatrix} N\sum_i x_i d_i - \sum_i x_i \sum_i d_i \\ \sum_i x_i^2 \sum_i d_i - \sum_i x_i \sum_i x_i d_i \end{bmatrix}
\end{aligned}
\tag{16}
$$

We can recognize the usual equation for the parameter of the linear regression.

In order to get the covariance matrix we have to compute the following expression:

$$
\begin{aligned}
\mathrm{Cov} &= \sqrt{\frac{1}{\det(\frac{\partial^2\chi}{\partial A_k \partial A_m})}\frac{\partial^2\chi}{\partial A_k \partial A_m}} \\
&= \sqrt{\frac{1}{\det(TC^{-1}T^t)}TC^{-1}T^t} \\
&= \sqrt{\frac{1}{\Delta}\begin{bmatrix} N & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}}
\end{aligned}
\tag{17}
$$