Mode of Multivariate Gaussian Distribution

Premmi

May 11, 2016

Proof. The pdf of $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{x}-\boldsymbol{\mu})\right\} \tag{1}$$

where $\mathbf{x}$ is d dimensional.

To find the mode i.e. the maximum of the Gaussian distribution, we differentiate the pdf with respect to $\mathbf{x}$ and equate it to 0 to find the critical point where the function is maximum or minimum and then we use the second derivative test to ascertain that the function is maximized at that point.*(The second derivative test for a function of more than one variable generalizes to a test based on the eigenvalues of the function's Hessian matrix at the critical point. Assuming that all the second order partial derivatives of the function are continuous on a neighbourhood of a critical point $\boldsymbol{x}$, then if the eigenvalues of the Hessian at $\boldsymbol{x}$ are all **positive**, then $\boldsymbol{x}$ is a local **minimum**, if the eigenvalues are all **negative**, then $\boldsymbol{x}$ is a local **maximum**, and if some are positive and some negative, then the point is a **saddle point**. If the Hessian matrix is **singular**, then the second derivative test is **inconclusive**.)*

Differentiating (1) with respect to $\mathbf{x}$,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{x}-\boldsymbol{\mu})\right\} \left\{-\boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{x}-\boldsymbol{\mu})\right\} \tag{2}$$

$$= -f(\mathbf{x})\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{x}-\boldsymbol{\mu}) \tag{3}$$

Equating (2) to zero, we get the critical point

$$\mathbf{x} = \boldsymbol{\mu} \tag{4}$$

To verify that the pdf is maximized at (4), we evaluate the Hessian matrix at $\boldsymbol{\mu}$ and check that it is negative definite.

Differentiating (3) with respect to $\mathbf{x}$,

$$\begin{aligned}
\frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^T} &= \frac{\partial}{\partial\mathbf{x}}\left(\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}^T}\right) &(5)\\
&= \frac{\partial}{\partial\mathbf{x}}\left(-f(\mathbf{x})\left(\mathbf{x}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{\mathbf{-1}}\right) &(6)\\
&= f(\mathbf{x})\boldsymbol{\Sigma}^{\mathbf{-1}}\left(\mathbf{x}-\boldsymbol{\mu}\right)\left(\mathbf{x}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{\mathbf{-1}} - f(\mathbf{x})\boldsymbol{\Sigma}^{\mathbf{-1}} &(7)\\
&= f(\mathbf{x})\left(\boldsymbol{\Sigma}^{\mathbf{-1}}\left(\mathbf{x}-\boldsymbol{\mu}\right)\left(\mathbf{x}-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{\mathbf{-1}} - \boldsymbol{\Sigma}^{\mathbf{-1}}\right) &(8)
\end{aligned}$$

Evaluating (8) at $\mathbf{x} = \boldsymbol{\mu}$ we get,

$$\left.\frac{\partial^2 f(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}\right|_{\mathbf{x}=\boldsymbol{\mu}} = -f(\boldsymbol{\mu})\boldsymbol{\Sigma}^{\mathbf{-1}} \tag{9}$$

We know that the covariance matrix $\boldsymbol{\Sigma}$ is positive definite, hence its inverse $\boldsymbol{\Sigma}^{\mathbf{-1}}$ is also positive definite.*(since if λ is an eigenvalue of $\boldsymbol{\Sigma}$, then $1/\lambda$ is the corresponding eigenvalue of $\boldsymbol{\Sigma}^{\mathbf{-1}}$. As λ is positive, so is $1/\lambda$ which implies that all the eigenvalues of $\boldsymbol{\Sigma}^{\mathbf{-1}}$ are positive.)*

We know that $f(\boldsymbol{\mu})$ is positive everywhere, since the pdf is positive, and hence $-\boldsymbol{\Sigma}^{\mathbf{-1}}$ is negative definite which proves that the pdf is maximized at $\mathbf{x} = \boldsymbol{\mu}$.*(if $\mathbf{Ax} = \lambda\mathbf{x}$, then $-\mathbf{Ax} = -\lambda\mathbf{x}$)*

Hence the mode of the multivariate Gaussian distribution is $\boldsymbol{\mu}$. $\square$

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning.* **Exercise 1.9**