Lecture 2: Gaussian Distributions

Given a continuous, random variable x which has a mean $\overline{x}$ and variance σ^2, a Gaussian probability distribution takes the form (Fig. 1):

$$P\{x\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\overline{x})^2}{2\sigma^2}\right\} \tag{1}$$

where σ is the standard deviation or the width of the Gaussian. We are interested in Gaussians because we shall assume that errors and uncertainties in data and in models follow this distribution.

The effect of increasing σ is to broaden and lower the peak height of the distribution, whereas changing $\overline{x}$ simply shifts the distribution along the x–axis.

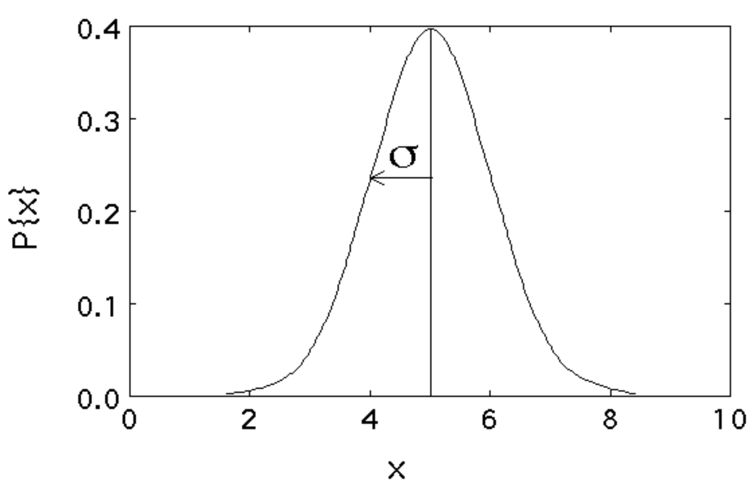

Fig. 1: One–dimensional Gaussian distribution with $\overline{x} = 5$ and $\sigma = 1$. Some 67% of the area under the curve lies between $\pm\sigma$, 95% between $\pm 2\sigma$ and 99% between $\pm 3\sigma$.

The factor $\frac{1}{\sigma\sqrt{2\pi}}$ arises from the normalisation of the distribution so that

$$\int_{-\infty}^{+\infty} P\{x\}\ dx = 1$$

i.e. the total probability, which is the area under the curve, is one.

Note that the term within the exponential in equation 1 can be differentiated twice with respect to x to obtain the variance:

$$\frac{\partial^2}{\partial x^2}\left[\frac{(x-\overline{x})^2}{2\sigma^2}\right] = \frac{1}{\sigma^2} \tag{2}$$

Two–Dimensional Gaussian Distribution

A two–dimensional Gaussian involves two random variables, x_1 and x_2 with mean values $\overline{x}_1$ and $\overline{x}_2$ (Fig. 2). A particular combination of x_1 and x_2 can be represented as a column vector:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \text{and} \qquad \mathbf{x}^T = (\,x_1, x_2\,)$$

where the superscript T indicates a transpose of the column vector into a row vector. Note that vectors are represented using bold font.

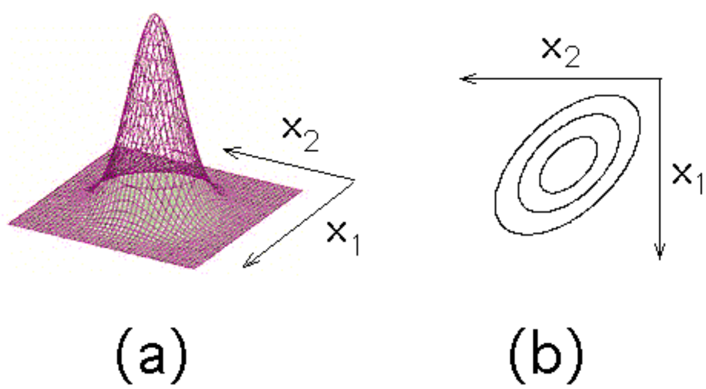

Fig. 2: (a) Two–dimensional Gaussian distribution. (b) Constant probability contour plot representing 2–D Gaussian.

The entire dataset consisting of n targets and n corresponding–inputs; can also be represented as matrices:

$$\mathbf{T} = \begin{pmatrix} t^{(1)} \\ t^{(2)} \\ \vdots \\ t^{(n)} \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} \\ x_2^{(1)} & x_2^{(2)} \\ \vdots & \vdots \\ x_n^{(1)} & x_n^{(2)} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

The mean values corresponding to each variable can be written as a vector $\overline{\mathbf{x}} = (\overline{x}_1, \overline{x}_2)^T$. Each variable will have a variance σ_1^2 and σ_2^2. However, it is possible that the variables are related in some way, in which case there will be *covariances* σ_{12} and σ_{21} with $\sigma_{12} = \sigma_{21}$, all of which can be incorporated into a *variance–covariance* matrix:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{pmatrix} \tag{3}$$

The Gaussian is then given by

$$P\{\mathbf{x}\} = \underbrace{\frac{1}{\sqrt{(2\pi)^k |\mathbf{V}|}}}_{\text{normalisation factor}} \exp\left[-\frac{1}{2}(\mathbf{x}-\overline{\mathbf{x}})^T \mathbf{V}^{-1} (\mathbf{x}-\overline{\mathbf{x}})\right] \tag{4}$$

where k is the number of variables ($k = 2$ since we only have x_1 and x_2) and $|\mathbf{V}|$ is the determinant of $\mathbf{V}$.

The equivalent of the term

$$\left[\frac{(x-\overline{x})^2}{2\sigma^2}\right]$$

of the one–dimensional Gaussian (equation 1) is for a two–dimensional Gaussian given by (equation 4)

$$M = \frac{1}{2}(\mathbf{x}-\overline{\mathbf{x}})^T \mathbf{V}^{-1} (\mathbf{x}-\overline{\mathbf{x}})$$

It follows by analogy with equation 2, that

$$\nabla\nabla M = \mathbf{V}^{-1} \tag{5}$$

where the operator ∇ implies differentiation with respect to $\mathbf{x}$.

Finally, it is worth noting that the product of two Gaussians is also a Gaussian.

More about the variance matrix

The variance matrix in equation 3 is symmetrical, which means that it can be expressed in another set of axes (known as the *principal axes*) such that all the off–diagonal terms become zero. This is evident from Fig. 3, where with respect to the dashed–axes, the off–diagonal terms must clearly be zero.

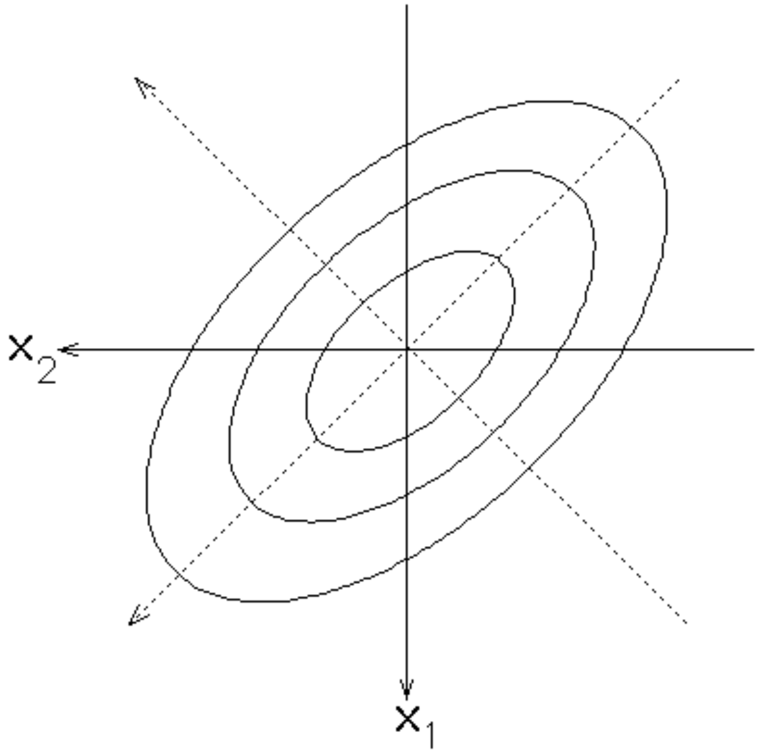

Fig. 3: Contours representing 2–D Gaussian distribution. The dashed axes are the principal axes.