



Figure 3.8 Illustration of a mixture of three Gaussians in a two-dimensional space. (a) Contours of constant density for each of the mixture components, in which the three components are denoted red, blue, and green, and the values of the mixing coefficients are shown below each component. (b) Contours of the marginal probability density $p(\mathbf{x})$ of the mixture distribution. (c) A surface plot of the distribution $p(\mathbf{x})$.

From the sum and product rules of probability, the marginal density can be written as

$$p(\mathbf{x}) = \sum_{k=1}^K p(k)p(\mathbf{x}|k), \quad (3.114)$$

which is equivalent to (3.111) in which we can view $\pi_k = p(k)$ as the prior probability of picking the k th component, and the density $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = p(\mathbf{x}|k)$ as the probability of \mathbf{x} conditioned on k . As we will see in later chapters, an important role is played by the corresponding posterior probabilities $p(k|\mathbf{x})$, which are also known as *responsibilities*. From Bayes' theorem, these are given by

$$\begin{aligned} \gamma_k(\mathbf{x}) &\equiv p(k|\mathbf{x}) \\ &= \frac{p(k)p(\mathbf{x}|k)}{\sum_l p(l)p(\mathbf{x}|l)} \\ &= \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_l \pi_l \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}. \end{aligned} \quad (3.115)$$

The form of the Gaussian mixture distribution is governed by the parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$, where we have used the notation $\boldsymbol{\pi} \equiv \{\pi_1, \dots, \pi_K\}$, $\boldsymbol{\mu} \equiv \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K\}$, and $\boldsymbol{\Sigma} \equiv \{\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K\}$. One way to set the values of these parameters is to use maximum likelihood. From (3.111), the log of the likelihood function is given by

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \quad (3.116)$$

where $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$. We immediately see that the situation is now much more complex than with a single Gaussian, due to the summation over k inside the log-