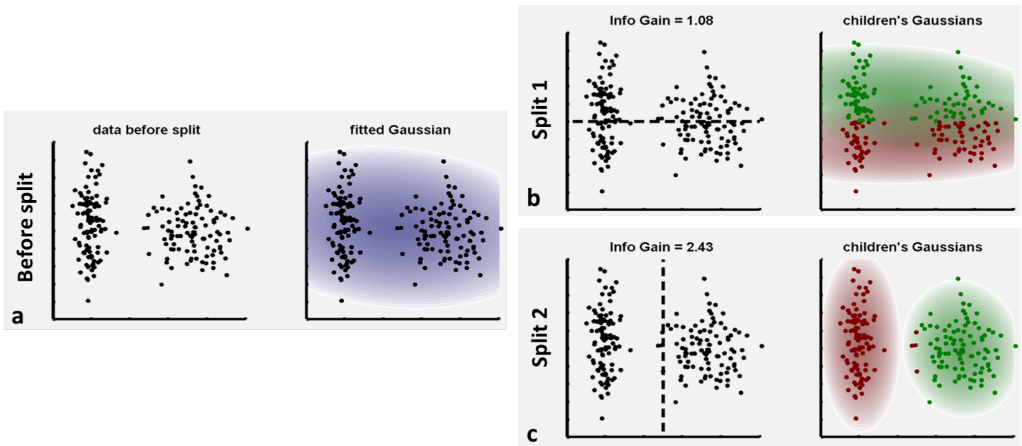

Fig. 2.4: **Information gain for continuous, parametric densities.** **(a)** Dataset $\mathcal{S}$ before a split. **(b)** After a horizontal split. **(c)** After a vertical split.

The previous example has focused on discrete, categorical distributions. But entropy and information gain can also be defined for continuous distributions. In fact, for instance, the differential entropy of a d-variate Gaussian density is defined as.

$$H(\mathcal{S}) = \frac{1}{2} \log \left((2\pi e)^d |\Lambda(\mathcal{S})| \right)$$

An example is shown in fig. 2.4. In fig. 2.4a we have a set $\mathcal{S}$ of *unlabelled* data points. Fitting a Gaussian to the entire initial set $\mathcal{S}$ produces the density shown in blue. Splitting the data horizontally (fig. 2.4b) produces two largely overlapping Gaussians (in red and green). The large overlap indicates a suboptimal separation and is associated with a relatively low information gain ($I = 1.08$). Splitting the data points vertically (fig. 2.4c) yields better separated, peakier Gaussians, with a correspondingly higher value of information gain ($I = 2.43$). The fact that the information gain measure can be defined flexibly, for discrete and continuous distributions, for supervised and unsupervised data is a useful property that is exploited here to construct a unified forest framework to address many diverse tasks.