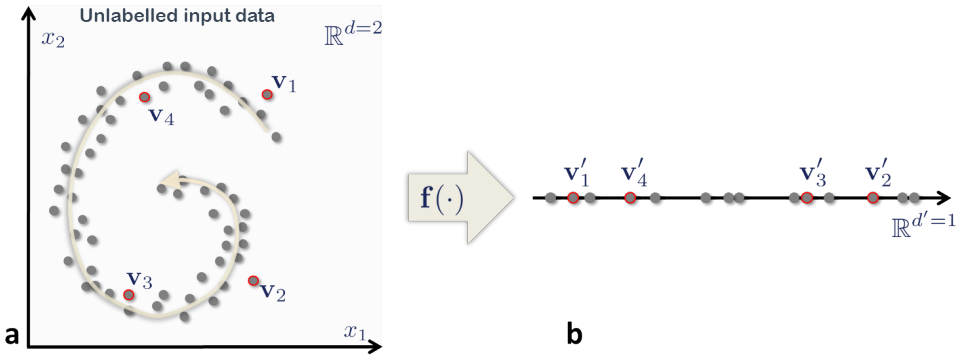

Fig. 6.1: **Manifold learning and dimensionality reduction. (a)** Input, unlabelled data points are denoted with circles. They live in a high-dimensional space (here $d = 2$ for illustration clarity). A red outline highlights some selected points of interest. **(b)** The target space is much lower dimensionality (here $d' = 1$ for illustration). Geodesic distances and ordering are preserved.

gain measure

$$\boldsymbol{\theta}_j^* = \arg\max_{\boldsymbol{\theta}_j \in \mathcal{T}_j} I_j$$

with I_j defined as for density forests:

$$I_j = \log(|\Lambda(\mathcal{S}_j)|) - \sum_{i \in \{L,R\}} \frac{|\mathcal{S}_j^i|}{|\mathcal{S}_j|} \log(|\Lambda(\mathcal{S}_j^i)|). \tag{6.1}$$

The previous chapter has discussed properties and advantages of (6.1).

The predictor model. Like in the density model the statistics of all training points arriving at each leaf node is summarized with a single multi-variate Gaussian:

$$p_t(\mathbf{v}) = \frac{\pi_{l(\mathbf{v})}}{Z_t} \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_{l(\mathbf{v})}, \Lambda_{l(\mathbf{v})}).$$

The affinity model. Unlike other tasks, in manifold learning we need to estimate some measure of similarity or distance between data