

by training with the squared error criterion

$$\|g(f(\tilde{\mathbf{x}})) - \mathbf{x}\|^2 \quad (14.16)$$

and corruption

$$C(\tilde{\mathbf{x}} = \tilde{\mathbf{x}}|\mathbf{x}) = \mathcal{N}(\tilde{\mathbf{x}}; \mu = \mathbf{x}, \Sigma = \sigma^2 I) \quad (14.17)$$

with noise variance  $\sigma^2$ . See figure 14.5 for an illustration of how this works.

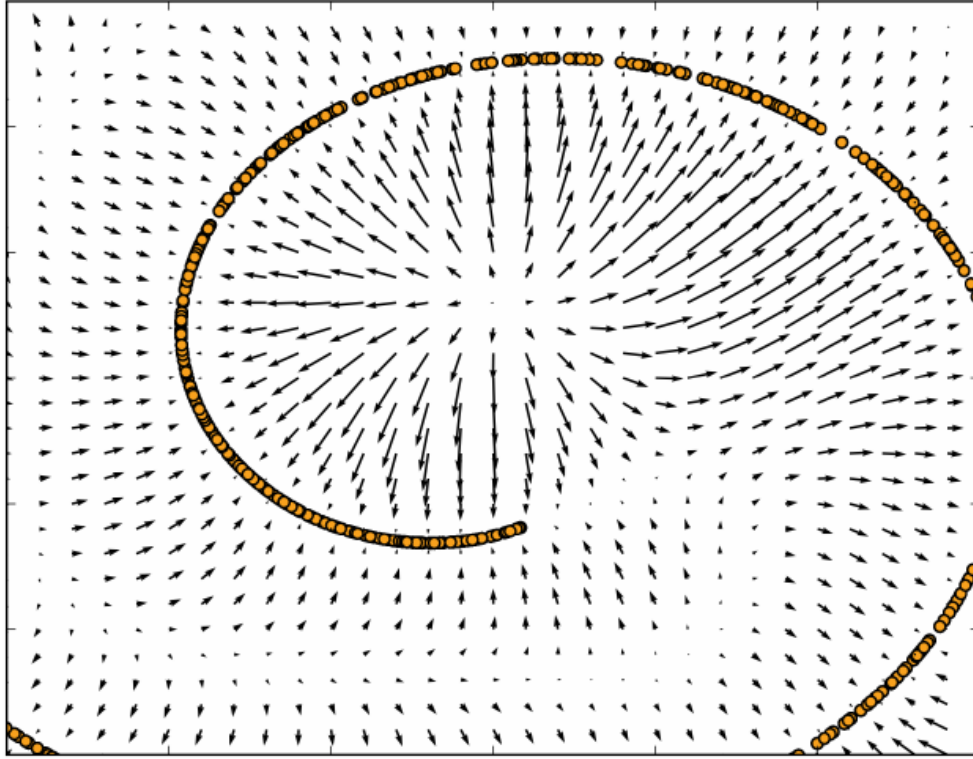


Figure 14.5: Vector field learned by a denoising autoencoder around a 1-D curved manifold near which the data concentrates in a 2-D space. Each arrow is proportional to the reconstruction minus input vector of the autoencoder and points towards higher probability according to the implicitly estimated probability distribution. The vector field has zeros at both maxima of the estimated density function (on the data manifolds) and at minima of that density function. For example, the spiral arm forms a one-dimensional manifold of local maxima that are connected to each other. Local minima appear near the middle of the gap between two arms. When the norm of reconstruction error (shown by the length of the arrows) is large, it means that probability can be significantly increased by moving in the direction of the arrow, and that is mostly the case in places of low probability. The autoencoder maps these low probability points to higher probability reconstructions. Where probability is maximal, the arrows shrink because the reconstruction becomes more accurate. Figure reproduced with permission from [Alain and Bengio \(2013\)](#).

In general, there is no guarantee that the reconstruction  $g(f(\mathbf{x}))$  minus the input  $\mathbf{x}$  corresponds to the gradient of any function, let alone to the score. That is