



Figure 10.9: An RNN that maps a fixed-length vector \mathbf{x} into a distribution over sequences \mathbf{Y} . This RNN is appropriate for tasks such as image captioning, where a single image is used as input to a model that then produces a sequence of words describing the image. Each element $\mathbf{y}^{(t)}$ of the observed output sequence serves both as input (for the current time step) and, during training, as target (for the previous time step).