



Figure 6.14 Illustration of three different contrastive learning paradigms. (a) The instance discrimination approach, where the positive pair is made up of the anchor and an augmented version of the same image. These are mapped to points in a normalized space that can be thought of as a unit hypersphere. The coloured arrows show that the loss encourages the representations of the positive pair to be closer together but pushes negative pairs further apart. (b) Supervised contrastive learning in which the positive pair consists of two different images from the same class. (c) The CLIP model in which the positive pair is made up of an image and an associated text snippet.

set with paired modalities can be used to learn representations. A comparison of the different contrastive learning methods we have discussed is shown in Figure 6.14.

6.3.6 General network architectures

So far, we have explored neural network architectures that are organized into a sequence of fully-connected layers. However, because there is a direct correspondence between a network diagram and its mathematical function, we can develop more general network mappings by considering more complex network diagrams. These must be restricted to a *feed-forward* architecture, in other words to one having no closed directed cycles, to ensure that the outputs are deterministic functions of the inputs. This is illustrated with a simple example in Figure 6.15. Each (hidden or output) unit in such a network computes a function given by

$$z_k = h \left(\sum_{j \in \mathcal{A}(k)} w_{kj} z_j + b_k \right) \quad (6.22)$$

where $\mathcal{A}(k)$ denotes the set of *ancestors* of node k , in other words the set of units that send connections to unit k , and b_k denotes the associated bias parameter. For a given set of values applied to the inputs of the network, successive application of (6.22) allows the activations of all units in the network to be evaluated including those of the output units.