



Generative Adversarial Networks (GAN)

Generative Adversarial Networks fall between supervised learning based on the provision of input data, the corresponding desired output of which is known in order to estimate the input-output relationship beyond the samples provided for training, and unsupervised learning.

When only input data is available, to discover certain structures within the data (e.g. the number of parameters that characterise it), there are many other paradigms, for example **semi-supervised** where we mix data where we know the desired output and others do not, in order to mix the two approaches.

Another so-called **self-supervised** paradigm consists of, from input data, finding an external mechanism to generate the corresponding outputs. This is to save the enormous human effort of entering for each input the desired output, e.g. labelling images by hand, pixel by pixel if necessary (if one wants to find where the cat is in an image). For example, to learn to automatically colourise images, one can start with colour images, reduce them to black and white images, and then train the mechanism by providing it with the desired colour images, here known without the need to reconstruct them for each black and white image.

This also works for learning the relative position of elements in an image that is being cropped, or temporal consistency in a video. But it doesn't work for everything. It works whenever you find a trick to automatically generate the desired inputs and output from the data. **This is sort of unsupervised learning that auto-generates data for a supervised learning paradigm.**

