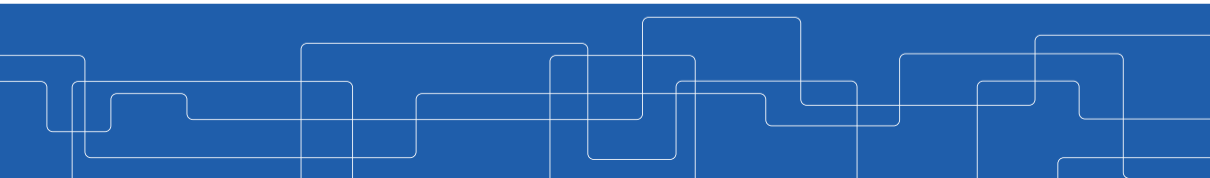




Variational Auto Encoders

Lukas Käll
lukask@kth.se

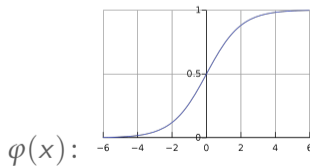
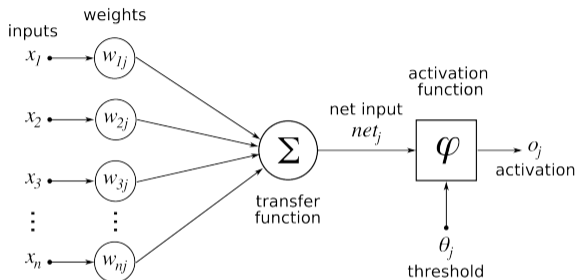


Artificial Neuron

A mathematical construction that takes one or more inputs and sums them to produce an output, often through a non-linear function, φ .

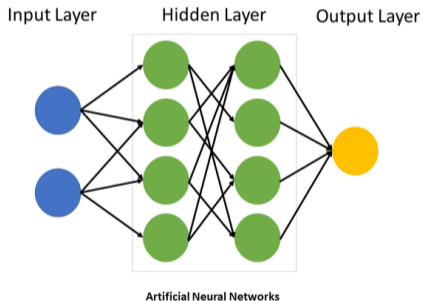
$$o_j = \varphi\left(\sum_{i=1}^n x_i w_{ij}\right)$$

$$\varphi(x) = \frac{1}{1 + e^{-\beta x}}$$



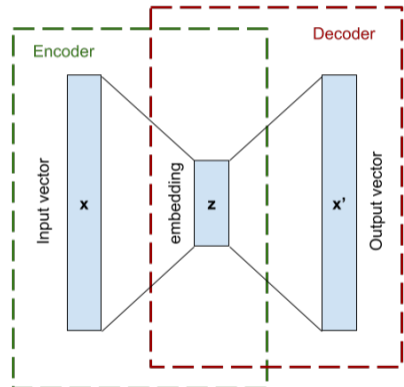
Artificial Neural Network

Artificial Neurons can be combined into an Artificial Neural Network (ANN). Such networks consist of an input layer, with the same dimensions as your input vector, an output layer, with the size of the output vector, and a set of hidden layers of unspecified dimensions. One can train an ANN to mimic almost any mapping between an input and output vector, $f : x \mapsto o$, using supervised training.



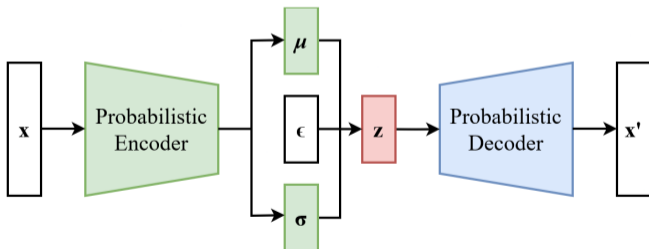
Auto encoder

If we don't have labeled data, we can implement an autoencoder, a combination of two ANNs, an encoder, that maps an input vector x , to a lower dimensional vector h , a decoder that converts a vector h to a vector x' of the same dimension as x . We can train the network with the same output x' as input x . Once trained we can use the h vectors, the so-called embeddings, as a representation of x . They contain the information needed to recreate the full x' through the decoder.



Variational Auto encoder

A variational autoencoder (VAE) is an autoencoder whose training is regularised to avoid overfitting and ensure that the latent space has good properties that enable generative process. A encoder of a VAE has a probabilistic interpretation, as it generates a probability distribution, from which we can sample, $z \sim Pr(z|x)$, and a decoder that gives $x' = d(z)$





Thanks!

Images from Wikipedia under CC-BY licence.