

2019/05/25 8:36:30 AM

Adds a tunable hyper parameter Beta as a weight for the 2nd term from the standard variational auto encoder optimization function;

the point is apparently to discover interpretable factorised latent representations from raw image data - 2 requirements are to be satisfied here: independence and interpretability;

- many simpler methods like ICA and PCA might give independent components but not necessarily interpretable ones

Some hacky theoretical justification is given for this in the form of having to come up with a representation/distribution that satisfies the 2 criteria of statistical independence and maximizing the likelihood of the observed data being generated from the target distribution

Apparently, such a distribution can be obtained by trying to constrain or match it to a prior of an isotropic unit Gaussian distribution which in turn is achieved by using the KKT conditions for constrained optimization

Also introduces a protocol for quantitative evaluation of the disentangling capability of a representation using simulated data:

- Generate a set of images by fixing only one latent factor while randomly varying all others
- assume that if the representation is able to correctly identify/disentangle the fixed latent factor, the variance of the corresponding features would be less than that of the remaining factors
- therefore, a simple linear classifier would be able to take this representation (encoder output) and predict the fixed latent factor correctly from it
- pairwise absolute differences are computed between the reconstructed images corresponding to a given fixed leading factor and then it is the average of these differences that is subjected to the classification

More disentangled representations apparently lead to poorer reconstruction quality with blurry results while it is the most entangled representations that lead to the best reconstructions

- Optimizing Reconstruction quality is therefore not the best way to learn the most disentangled representation though it should be possible

to get values of Beta that provide a good balance between
disentanglement capacity of the representation and the corresponding
reconstruction quality