Progressive Learning and Disentanglement of Hierarchical Representations

1. Core Idea

Starting Small:

2. Model OV

Hierarchical generative model with z decomposed to $z_1, z_2, ..., z_L$ different abstraction levels,

$$p(x,z) = p(x|z_1, z_2, ..., z_L)\prod_{l=1}^{L} p(z_l)$$

Inference model,

$$q(z_1, z_2, ..., z_L|x) = \prod_{l=1}^{L} q(z_l|h_l(x))$$

Optimize using β-VAE objective. Prior $p(z)$ $N(0, I)$.

3. Learning the Model

Progressively learn from highest ($l = L$) to lowest ($l = 1$) level of abstraction. Inference model at step s=0,

$$z_L \; N(\mu_L(h_L), \sigma_L(h_L))$$
$$h_l = f_l^e(h_{l-1})$$

where $h_0 = x$.

Generative model at step s=0,

$$g_L = f_L^d(z_L)$$
$$g_l = f_l^d(g_{l+1})$$
$$x = D(x; f_0^d(g_0))$$

4. Disentanglement Metric

MIG measures the split of factors into multiple dimensions. But the score will still be good if multiple factors are in the same dimension (but not split).

$$\frac{1}{J}\sum_{1}^{J}(I_{norm}(z_j; v_{k^j}) - \max_{k \neq k^{(j)}} I_{norm}(z_j; v_k))$$

where v is ground truth factors.