4.3 $$\mathcal{L}_*(h_\psi, g_\phi, D) = \frac{1}{N}\sum_{i=1}^{N} -\log\left(\frac{\exp\left(h_\psi(x_i, x_i')^T \cdot g_\phi(x_i')\right)}{\frac{1}{M}\sum_{j=1}^{M} \exp\left(h_\psi(x_i, x_i')^T \cdot g_\phi(x_j)\right)}\right)$$

In this formulation, we recover similar terms:

$$\underline{\mathcal{L}_{align}} = \frac{-1}{N}\sum_{i=1}^{N} h_\psi(x_i, x_i') \cdot g_\phi(x_i')$$

which aims to align positive representations.

$$\underline{\mathcal{L}_{unif}} = \frac{1}{N}\sum_{i=1}^{N} \log \frac{1}{M}\sum_{j=1}^{M} e^{h_\psi(x_i, x_i')^T \cdot g_\phi(x_j)}$$

which promotes uniform distribution of negative representations.

$\log(M)$: rescaling constant.

4.3

EBM with L_{contr}:

This design is straight forward, as the model produces embeddings independently. However, representations may not be as rich since they both use one data modality.

EBM with L^*:

This version has the advantage of potentially capturing richer representations with h_ψ, since it uses two different modalities and may capture features f_θ and g_ϕ can't. However, the trade off is that h_ψ will likely need higher model capacity and more data in order to do so.