TTIC 31230 Fundamentals of Deep Learning

Contrastive Learning Problems

Problem 1: Upper Bounding $H(y)$

We consider a population distribution Pop on a set of observable values y and a stochastic encoder defining a conditional distribution $P_\Psi(z|y)$. We assume that we can sample from $P_\Psi(z|y)$ and that for any given z and y we can compute $P_\Psi(z|y)$. The population and the encoder define a joint distribution $P_{\mathrm{Pop},\Psi}(y,z)$ where y is drawn from the population and z is drawn from $P_\Psi(z|y)$. All probabilities and information-theoretic quantities in this problem refer to this joint distribution.

We will use the fact that mutual information satisfies

$$I(y,z) = H(y) - H(y|z) = H(z) - H(z|y)$$

which implies

$$H(y) = H(z) - H(z|y) + H(y|z) \tag{1}$$

(a) Rewrite (1) in terms of expectations over $y \sim \mathrm{Pop}$ and $z \sim P_\Psi(z|y)$ of quantities defined on $\mathrm{Pop}(y)$, $P_{\mathrm{Pop},\Psi}(z)$, $P_{\mathrm{Pop},\Psi}(z|y)$ and $P_{\mathrm{Pop},\Psi}(y|z)$.

(b) Which of the terms in (1) can be directly esimated by simply sampling $y \sim \mathrm{Pop}$ and $z \sim P_\Psi(z|y)$.

(c) Recall that the cross-entropy $H(P,Q)$ is defined to be $E_{x\sim P}\ -\ln Q(x)$ and that $H(P) \leq H(P,Q)$ for any Q. Let $P_\Phi(z)$ and $P_\Theta(y|z)$ be two additional models and consider the cross entropies $H(P_{\mathrm{Pop},\Psi}(z), P_\Phi(z))$ and $H(P_{\mathrm{Pop},\Psi}(y|z), P_\Theta(y|z))$. Using the fact that cross-entropies upper bound entropies give an upper bound on $H(y)$ derived from (1) by replacing entropies by cross-entropies to these models. Express your upper bound as an expectation over sampling.

(d) Which terms in you solution to (c) can be estimated directly by sampling.

(e) Consider minimizing the upper bound on $H(y)$ given in your solution to (c). How is this related the VAE training objective?