TTIC 31230 Fundamentals of Deep Learning
Exam 1: 10% of class grade

In all problems we assume that all probability distributions $P(x)$ are discrete so that we have $\sum_x \; P(x) = 1$.

Problem 1 (25 pts): We define conditional entropy $H(y|x)$ as follows

$$H(y|x) = E_{x,y} \; - \log P(y|x).$$

Given a distribution $P(x, y)$ show

$$H(P) = H(x) + H(y|x).$$

Problem 2 (25 pts) Consider a joint distribution $P(x, y)$ on discrete random variables x and y. We define the marginal distributions $P(x)$ and $P(y)$ as follows.

$$\begin{aligned} P(x) &= \sum_y P(x,y) \\ P(y) &= \sum_x P(x,y) \end{aligned}$$

Let $Q(x, y)$ be defined to be the product of marginals.

$$Q(x, y) = P(x)P(y).$$

Derive the following equalities.

$$KL(P(x, y), Q(x, y)) = H(y) - H(y|x) = H(x) - H(x|y)$$

The above quantity is called the mutual information between x and y, written $I(x, y)$. Explain why this quantity is always non-negative.

Problem 3 (25 pts): Consider two (possibly unrelated) distributions $P(z, x)$ and $Q(z|x)$. Show that for any specific value of x we have

$$E_{z\sim Q(z|x)} \; \ln P(z, x) = \ln \; P(x) - H(Q(z|x)) - KL(Q(z|x), P(z|x)).$$

Show that this implies

$$\ln P(x) \geq \left(E_{z\sim Q(z|x)} \; \ln \; P(z, x)\right) + H(Q(z|x))$$

This last inequality is called the evidence lower bound (the ELBO). This terminology comes from viewing an observed variable x as evidence for a latent variable z. The ELBO is the core of expectation maximization (EM) and variational auto encoders (VAEs).

Problem 4 (25 pts) For three distributions P, Q and G show the following equality.

$$KL(P,Q) = \left(E_{x\sim P}\ \log\frac{G(x)}{Q(x)}\right) + KL(P,G)$$

Explain why this implies

$$KL(P,Q) \geq\ E_{x\sim P}\ \log\frac{G(x)}{Q(x)}$$

Next define

$$\begin{aligned} G(x) &= \frac{1}{Z}\,Q(x)e^{s(x)} \\ Z &= \sum_x Q(x)e^{s(x)} \end{aligned}$$

Show that this definition of $G(x)$ gives

$$KL(P,Q) \geq\ E_{x\sim P}\ s(x) - \log E_{x\sim Q}\ e^{s(x)}$$

This is the Donsker-Varadhan lower bound on KL-divergence.