TTIC 31230 Fundamentals of Deep Learning

Problems for Latent Variables, EM, the ELBO and VAEs.

Problem 1. Assume a population distribution on pairs (x, y) where x is an image and y is a semantic segmentation of x. The semantic segmentation y assigns a class label $y[i] \in \mathcal{C}$ to each pixel i. A sparse segmentation $\tilde{y}$ assigns a class label or a blank to each pixel with $\tilde{s}[i] \in \mathcal{C} \cup \{\perp\}$. We expect most pixels to be blank. Suppose we want to learn a compression algorithm $\tilde{y}_\Phi(x, y)$ that compresses a given semantic segmentation y into a sparse semantic segmentation $\tilde{y}_\Phi(x, y)$ and a decompression algorithm $y_\Phi(x, \tilde{y})$ that decompresses the sparse segmentation into a full segmentation. For two segmentation y and y' let $\mathrm{Dist}(y, y')$ be the Hamming distance between them — the number of pixels on which they disagee.

(a) Assume a representation of a spare coding as a list of pairs (i, c) with i a pixel and c a semantic category. Assume there are P pixels and k semantic categories. Define a probability distribution $P(L)$ for lists of this form where $\ln P(L)$ is proportional to the length of the list L.

(b) Use your answer to (a) to define a rate-distortion objective function for optimizing Φ.

(c) What advantage does this rate-distortion autoencoder architecture have over a graphical model?

(d) What are the challenges in training this rate-distortion autoencoder.

Problem 2. Consider a conditional latent variable model satisfying.

$$P_\Phi(y|x) = \sum_z P_\Phi(z|x) P_\Phi(y|z, x)$$

(a) Show

$$E_{(x,y)\sim\mathrm{Pop}} - \ln P_\Phi(y|x) = \inf_Q \; E_{(x,y)\sim\mathrm{Pop},\; z\sim Q(z|y,x)} - \ln \frac{P_\Phi(z, y|x)}{Q(z|y, x)}.$$

The expression on the right hand is the negative of the ELBO — the ELBO loss to be minimized. Hint: Since all probabilities involved are conditioned on x, we can essentially ignore x.

(b) Consider the negative ELBO optimization problem

$$\Phi^*, \Psi^* = \operatorname*{argmin}_{\Phi,\Psi} \; E_{y\sim\mathrm{Pop},\; z\sim p_\Psi(z|y,x)} - \ln \frac{P_\Phi(z, y|x)}{P_\Psi(z|y, x)}$$

What is the motivation for the ELBO optimization over the more direct optimization problem

$$\Phi^* = \operatorname*{argmin}_{\Phi} \; E_{y\sim \mathrm{Pop},\, z\sim p_\Phi(z|y,x)} - \ln P_\Phi(y|z,x) \quad ?$$