TTIC 31230 Fundamentals of Deep Learning

Quiz 4

Problem 1 (25 points) This problem is on Pseudolikelihood. Consider a graphical model with N nodes numbered 1 through N and where each node can take on one of the values 0 or 1. We let $\hat{x}$ be an assignment of a value to every node. We define the score of $\hat{x}$ by

$$s(\hat{x}) = \sum_{i=1}^{N-1} \mathbf{1}[\hat{x}[i] = \hat{x}[i+1]]$$

where

$$\mathbf{1}[S] = \begin{cases} 1 \text{ if statement } S \text{ is true} \\ 0 \text{ otherwise} \end{cases}$$

The probability distribution over assignments is defined by a softmax.

$$Q_s(\hat{x}) = \operatorname*{softmax}_{\hat{x}} s(\hat{x})$$

What is the **Pseudoliklihood** of the all ones assignment?

Solution:

$$\tilde{P}_s(\hat{x}) = \Pi_i P_s(\hat{x}[i] \mid \hat{x}/i)$$

where $\hat{x}/i$ consists of all components of $\hat{x}$ other than i. In a graphical model $P_s(\hat{x}[i] \mid \hat{x}/i)$ is determined by the neighbors of i and we can consider only how a value is scored against it neighbors. For $\hat{x}$ equal to all ones we have

$$\begin{aligned} s(\hat{x}) &= N-1 \\ s(\hat{x}[i=0]) &= \begin{cases} N-3 & \text{for } 1 < i < N \\ N-2 & \text{for } i = 1 \text{ or } i = N \end{cases} \end{aligned}$$

For $1 < i < N$ we get

$$\begin{aligned} Q_s(\hat{x}[i=1] \mid \hat{x}/i) &= \frac{e^{N-1}}{e^{N-1} + e^{N-3}} \\ &= \frac{1}{1+e^{-2}} \end{aligned}$$

and for $i = 1$ or $i = N$ we get

$$Q_s(\hat{x}[i=1] \mid \hat{x}/i) = \frac{1}{1+e^{-1}}$$

This gives

$$\tilde{Q}(\hat{x}) = (1 + e^{-1})^{-2}(1 + e^{-2})^{-(N-2)}$$

Problem 2. (25 points). Consider a rate-distortion autoencoder.

$$\Phi^* = \operatorname*{argmin}_{\Phi} \; I_\Phi(y, z) + \lambda E_{y \sim \text{pop},\; z \sim p_\Phi(z|y)} \; \text{Dist}(y, y_\Phi(z)).$$

Here $I_\Phi(y, z)$ is defined by the distribution where we draw y from pop and z from $P_\Phi(z|y)$ and where the mutual information is defined by

$$I(x, y) = E_{x,y} \; \ln \frac{p(x, y)}{p(x)p(y)} = KL(p(x, y), p(x)p(y))$$

The distribution $p_\Phi(z|y)$ is typically defined by $z = z_\Phi(y) + \epsilon$ for some form of random noise ϵ.

(a) Starting from the definition of $I(y, z)$ given above, show

$$I(y, z) = E_y \; KL(p(z|y), p(z))$$

where $p_\Phi(z) = \sum_y \text{pop}(y) P_\Phi(z|y)$.

Solution:

$$\begin{aligned}
I(z, y) &= KL(p(z, y), p(z)p(p)) \\
&= E_{z,y} \; \ln \frac{p(z, y)}{p(z)p(y)} \\
&= E_{z,y} \; \ln \frac{p(y)p(z|y)}{p(y)p(z)} \\
&= E_y \left(E_{z \sim p(z|y)} \; \ln \frac{p(z|y)}{p(z)} \right) \\
&= E_y \; KL(p(z|y), p(z))
\end{aligned}$$

(b) Show the variational equation

$$I(y, z) = \inf_q \; E_{y \sim \text{pop}} KL(p_\Phi(z|y), q(z)).$$

Hint: It suffices to show

$$I(y, z) \leq \; E_{y \sim \text{pop}} KL(p_\Phi(z|y), q(z))$$

and that there exists a q achieving equality.

Solution:

$$
\begin{aligned}
& I_\Phi(y,z) \\
= \;& E_{y\sim\mathrm{pop}}\, KL(p_\Phi(z|y), p_\Phi(z)) \\
= \;& E_{y,z\sim P_\Phi(z|y)} \left(\ln\frac{p_\Phi(z|y)}{q(z)} + \ln\frac{q(z)}{p_\Phi(z)}\right) \\
= \;& E_{y\sim\mathrm{pop}}\, KL(p_\Phi(z|y), q(z)) + \left(E_{y\sim\mathrm{pop},\; z\sim p_\Phi(z|y)}\; \ln\frac{q(z)}{p_\Phi(z)}\right) \\
= \;& E_y\; KL(p_\Phi(z|y), q(z)) + E_{z\sim p_\Phi(z)}\; \ln\frac{q(z)}{p_\Phi(z)} \\
= \;& E_y\; KL(p_\Phi(z|y), q(z)) - KL(p_\Phi(z), q(z)) \\
\leq \;& E_{y\sim\mathrm{pop}}\, KL(p_\Phi(z|y), q(z))
\end{aligned}
$$

From part (a) equality is achieved when $q(z) = p_\Phi(z)$.

(c) Based on the result from part (b) rewrite the definition of rate-distortion autoencoder to be a minimization over two models Φ and Ψ which gives the same meaning for Φ assuming universality for Ψ.

Solution:

$$\Phi^*, \Psi^* = \underset{\Phi,\Psi}{\mathrm{argmin}}\, E_{y\sim\mathrm{pop},z\sim P_\Phi(z|y)}\; \ln\frac{p_\Phi(z|y)}{p_\Psi(z)} + \lambda\, \mathrm{Dist}(y, y_\Phi(z)).$$

Problem 3. (25 points)

This problem is on the instability of adversarial objectives. Consider the following adversarial objective where x and y are scalars (real numbers).

$$\max_x\; \min_y\; xy$$

(a) Write the differential equation for gradient flow of this adversarial objective.

Solution:

$$\frac{dx}{dt} = y$$

$$\frac{dy}{dt} = -x$$

(b) Give a general solution to your differential equation. (Hint: It goes in a circle). You solution should have parameters allowing for any given initial value of x and y.

Solution:

$$x = r_0 \sin(t + \Theta_0)$$

$$y = r_0 \cos(t + \Theta_0)$$

Problem 4. (25 points)

This problem is on contrastive GANs.

Define the distribution $P \hookrightarrow Q^k$ to be the result of drawing one "positive" from P and k IID "negatives" from Q; then inserting the positive at a random position among the negatives; and returning $(i, y_1, \ldots, y_{N+1})$ where i is the index of the positive.

Consider a contrastive GAN.

$$\Phi^* = \operatorname*{argmin}_{\Phi} \max_{\Psi} \; E_{(i,y_1,\ldots,y_{N+1})\sim(\mathrm{pop}\hookrightarrow p_\Phi^k)} \ln p_\Psi(i|y_1,\ldots,y_{N+1}) \quad (1)$$

Here the objective function has been negated so that we have a min-max problem rather than a max-min problem.

(a) Convert the above equation to a conditional contrastive GAN for modeling $p_\Phi(y|x)$ rather than $p_\Phi(y)$ trained by the above equation. Your solution should have the property that the optimum assuming universality is the true conditional distribution $\mathrm{pop}(y|x)$.

Solution:

$$\Phi^* = \operatorname*{argmin}_{\Phi} \max_{\Psi} E_{x\sim\mathrm{pop}} \;\; E_{(i,y_1,\ldots,y_{N+1})\sim(\mathrm{pop}(y|x)\hookrightarrow p_\Phi(y|x)^k)} \ln p_\Psi(i|y_1,\ldots,y_{N+1},x)$$

(b) It is shown in the slides that if we restrict $p_\Psi(i|y_1,\ldots,y_{N+1})$ to be computed from a softmax of a score $s_\Phi(y)$, and assume universality, we have

$$p_{\Psi^*}(i|y_1,\ldots,y_{N+1}) = \operatorname*{softmax}_i \; \ln \frac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}$$

Given this property of Ψ^*, show

$$\begin{array}{rl} & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln p_{\Psi^*}(i|y_1,\ldots,y_{N+1}) \\ \\ \leq & \dfrac{N}{N+1}(KL(\mathrm{pop},p_\Phi)+KL(p_\Phi,\mathrm{pop})) + \ln \dfrac{1}{N+1} \end{array}$$

Note that if $p_\Phi = \mathrm{pop}$ then $P(i|y_1,\ldots,y_{N+1}) = \frac{1}{N+1}$ in which case this bound is tight.

Solution:

$$\begin{array}{rl} & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln p_{\Psi^*}(i|y_1,\ldots,y_{N+1}) \\ \\ = & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln \left(\operatorname*{softmax}_i \; \ln \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}\right)[i] \\ \\ = & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\left(\displaystyle\sum_i \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}\right) \\ \\ = & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\left(\dfrac{1}{N+1}\displaystyle\sum_i \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)}\right) - \ln\,(N+1) \\ \\ \leq & E_{(i,y_1,\ldots,y_{N+1})\sim \mathrm{pop}\hookrightarrow p_\Phi^N} \;\ln \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \dfrac{1}{N+1}\displaystyle\sum_i \ln \dfrac{\mathrm{pop}(y_i)}{p_\Phi(y_i)} - \ln\,(N+1) \\ \\ = & \dfrac{N}{N+1}E_{y\sim\mathrm{pop}} \ln \dfrac{\mathrm{pop}(y)}{p_\Phi(y)} - \dfrac{N}{N+1}E_{y\sim p_\Phi} \ln \dfrac{\mathrm{pop}(y)}{p_\Phi(y)} - \ln\,(N+1) \\ \\ = & \dfrac{N}{N+1}(KL(\mathrm{pop},p_\Phi)+KL(p_\Phi,\mathrm{pop})) - \ln\,(N+1) \end{array}$$