

CS6202: Problem Set 2

Question 1

Give and derive the formula for the probability density function of the last iterate of Inverse Autoregressive Flow (IAF) $\log q(\mathbf{z}_T|\mathbf{x})$ in Variational Autoencoder. Assume that our base distribution $q(\mathbf{z}_0|\mathbf{x})$ is an isotropic Gaussian with parameter μ_0 and σ_0 . This is similar to proof the Equation 11 in Inverse Autoregressive Flow paper.

Solution 1

Remember that in Normalizing Flows the probability density function of the final iterate is given as:

$$\log q(\mathbf{z}_T|\mathbf{x}) = \log q(\mathbf{z}_0|\mathbf{x}) - \sum_{t=1}^T \log \det \left| \frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}} \right| \quad (1)$$

Knowing that the base distribution $\log q(\mathbf{z}_0|\mathbf{x})$ is an isotropic Gaussian and log-determinant of the Jacobian in IAF is given as follow:

$$\log \det \left| \frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}} \right| = \sum_{i=1}^D -\log \sigma_i(\mathbf{z}_t) \quad (2)$$

$$= \sum_{i=1}^D -\log \sigma_{t,i} \quad (3)$$

We use $\sigma_{t,i} = \sigma_i(\mathbf{z}_t)$ to make the notation of dimension i of σ at iteration t shorter. We can compute the probability density function of the final iterate as:

$$\begin{aligned} \log q(\mathbf{z}_T|\mathbf{x}) &= \log q(\mathbf{z}_0|\mathbf{x}) - \sum_{t=1}^T \log \det \left| \frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}} \right| \\ &= \sum_{i=1}^D \left(\log \left[\frac{1}{\sqrt{2\pi}\sigma_{0,i}} \right] - \frac{(\mathbf{z}_0 - \mu_{0,i})^2}{2\sigma_{0,i}^2} \right) - \sum_{t=1}^T \log \det \left| \frac{d\mathbf{z}_t}{d\mathbf{z}_{t-1}} \right| \quad \dots \text{log of isotropic Gaussian} \\ &= \sum_{i=1}^D \log \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_{0,i}} \right] - \frac{(\mathbf{z}_0 - \mu_{0,i})^2}{2\sigma_{0,i}^2} - \sum_{t=1}^T -\log \sigma_{t,i} \quad \dots \text{applying Equation 3} \\ &= \sum_{i=1}^D -\frac{1}{2} \log(2\pi) - \log \sigma_{0,i} - \frac{1}{2} \epsilon_i^2 - \sum_{t=1}^T -\log \sigma_{t,i} \quad \dots \text{by } \epsilon = \frac{\mathbf{z}_0 - \mu(\mathbf{z}_0)}{\sigma(\mathbf{z}_0)} \\ &= -\sum_{i=1}^D \left(\frac{1}{2} \log(2\pi) + \frac{1}{2} \epsilon_i^2 + \sum_{t=0}^T \log \sigma_{t,i} \right) \quad \dots \text{merge the sum of } \sigma \text{ and take out negative.} \end{aligned}$$

Question 2

In practical applications of normalizing flows, present networks are tending to get deeper, i.e, more stacked

layers. This induces numerical stability issues. One of the ways to address this is by adding *Batch-normalization* layers. The forward pass of a typical batch-normalization layer is as follows:

$$y \mapsto \frac{x - \tilde{\mu}}{\sqrt{\tilde{\sigma}^2 + \epsilon}} \quad (4)$$

Derive the log determinant of the Jacobian for this transformation.

Solution 2

The log determinant of the Jacobian: $\log \left| \det \frac{\partial y}{\partial x} \right|$, can be calculated as follows:

$$\begin{aligned} \log \left| \det \frac{\partial y}{\partial x} \right| &= \log \left| \det \frac{\partial}{\partial x} \left(\frac{x - \tilde{\mu}}{\sqrt{\tilde{\sigma}^2 + \epsilon}} \right) \right| \\ &= \log \left| \det \left(\frac{\partial}{\partial x} \left(\frac{x}{\sqrt{\tilde{\sigma}^2 + \epsilon}} \right) - \frac{\partial}{\partial x} \left(\frac{\tilde{\mu}}{\sqrt{\tilde{\sigma}^2 + \epsilon}} \right) \right) \right| \\ &= \log \left| \det \left(\frac{1}{\sqrt{\tilde{\sigma}^2 + \epsilon}} \right) \right| \\ &= \sum_i \log \left| \frac{1}{\sqrt{\tilde{\sigma}_i^2 + \epsilon}} \right| \end{aligned}$$