Imperial College London

Department of Computing

Tutorial 5 - VAEs and GANs

1 Question 1

As described in the introduction, the generator and discriminator networks can be trained with the following objective function:

$$\min_{\theta} \max_{\phi} \mathcal{L} = \mathop{\mathbb{E}}_{\mathbf{x} \sim p_r(\mathbf{x})} [\log D_{\phi}(\mathbf{x})] + \mathop{\mathbb{E}}_{\mathbf{z} \sim q(\mathbf{z})} [\log (1 - D_{\phi}(G_{\theta}(\mathbf{z}))]$$

- **a:** Sketch the the function $\log(1-x)$ in the range $x \in [0,1]$ and argue with the help of your plot why the generator objective $\log(1 - D_{\phi}(G_{\theta}(\mathbf{z}))$ might lead to poor training when D rejects samples with high confidence, *ie.* $D(G_{\theta}(\mathbf{z}) \approx 0$ for any $\mathbf{z}$.

- **b:** Provided a fixed generator G, find the optimal discriminator $D^*(x)$.
  Hint1: treat the function $D(x)$ as your variable!
  Hint2: the objective can be formulated as

  $$\int p_r(\mathbf{x}) \log(D(\mathbf{x})) + p_g(\mathbf{x}) \log(1 - D(\mathbf{x})) d\mathbf{x}$$

- **c:** What is a divergence in the context of statistics? What is the formal definition?

- **d:** Assuming a perfect discriminator D^*, the objective $V(D^*, G)$ is

  $$\int p_r(\mathbf{x}) \log(D^*(\mathbf{x})) + p_g(\mathbf{x}) \log(1 - D^*(\mathbf{x})) d\mathbf{x}$$

  Derive that

  $$V(D^*, G) = 2D_{JS}(p_r \| p_g) - \log(4)$$

  Hint 1:

  $$D_{JS}(p_a \| p_b) = \frac{1}{2} D_{KL}(p_a \| \frac{p_a + p_b}{2}) + \frac{1}{2} D_{KL}(p_b \| \frac{p_a + p_b}{2})$$

  and

  $$D_{KL}(p_a \| p_b) = \int p_a \log(\frac{p_a}{p_b}) dx$$

  Hint 2: Recall the logarithm rules!

2 Question 2

- **a** For two different random events A and B, what is Bayes' Theorem?
- **b** Suppose you have a latent variable model (LVM) with model distribution $p_z\theta(\mathbf{x})$ and and latent variable $\mathbf{z}$ with distribution $p(\mathbf{z})$ as well as the conditional distribution $p_\theta(\mathbf{x}|\mathbf{z})$. You want to fit the model distribution $p_\theta(\mathbf{x})$ to the true (empirical) distribution $p(\mathbf{x})$. Define the maximum-likelihood objective in this context and argue why a direct optimisation of this objective is intractable if $p_\theta(\mathbf{x}|\mathbf{z})$ is represented by a complex neural network.
- **c** The KL-divergence for two multivariate gaussian distributions with means $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ and covariances Σ_0, Σ_1 has a closed form solution:

  $$D_{\text{KL}}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}\left(\text{tr}(\Sigma_1^{-1}\Sigma_0) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - K + \log\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right)\right)$$

  Show that given $\mathcal{N}_1$ is a standard normal distribution, *ie.* $\boldsymbol{\mu}_1 = \mathbf{0}$ and $\Sigma_1 = \boldsymbol{I}$ and $\mathcal{N}_0$ has diagonal covariance, the above equation reduces to:

  $$D_{\text{KL}}(\mathcal{N}_0 \| \mathcal{N}(\mathbf{0}, \boldsymbol{I})) = \frac{1}{2}\sum_{i+1}^{K} \sigma_i^2 + \mu_i^2 - 1 - \log(\sigma_i^2)$$