Department of Engineering/Informatics, King's College London
Pattern Recognition, Neural Networks and Deep Learning (7CCSMPNN)
Tutorial 6

Q1. Diminished gradient is an issue when training Generative Adversarial Networks (GANs). In the literature, when training the Generator, $\mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[-\log(D(G(\mathbf{z}))]$ is recommended to be an alternative cost function.

a. What is the advantage of using this alternative cost function over the original one, i.e., $\mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]$?

b. Write the Pseudo code for the training of GAN with this alternative cost function.

Q2. The training of Generative Adversarial Networks (GANs) can be formulated as an optimisation problem shown below:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))].$$

a. Find the optimal Discriminator $D(\mathbf{x})$, denoted as $D^*(\mathbf{x})$. Note that $\mathbf{x}$ denotes the sample taken by the Discriminator $D(\mathbf{x})$, which could represent the real or generated sample.

b. Find the optimal $V(D,G)$.

Q3. When training a Generative Adversarial Network (GAN), we consider a dataset of real samples denoted as $\mathbf{X}_{real} = \{\mathbf{x}_1, \mathbf{x}_2\}$ and the generated samples as $\mathbf{X}_{fake} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\tilde{\mathbf{x}}_1 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$, $\tilde{\mathbf{x}}_2 = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$.

The Discriminator is given as

$$D(\mathbf{x}) = \frac{1}{1 + e^{-(\theta_{d_1} x_1 - \theta_{d_2} x_2 - 2)}}$$

where θ_{d_1} and θ_{d_2} are parameters of Discriminator, and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Given $\theta_{d_1} = 0.1$ and $\theta_{d_2} = 0.2$. Each sample from the (real and fake) dataset has equal probability to be selected.

a. Given the datasets $\mathbf{X}_{real}$ and $\mathbf{X}_{fake}$, compute $V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\ln D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\ln(1 - D(G(\mathbf{z})))]$.

b. Assuming all real and fake samples are selected into the minibatch and $k = 1$ for GAN training, compute $\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\ln D\big(\mathbf{x}^{(i)}\big) + \ln\big(1 - D(G(\mathbf{z}^{(i)}))\big) \right]$ and determine the updated θ_{d_1} and θ_{d_2} for the next iteration using the learning rate $\eta = 0.02$.