Imperial College London

Department of Computing

Tutorial 5 - Solutions

1 Question 1

- **a** In order to update the generator G, gradients must be computed with respect

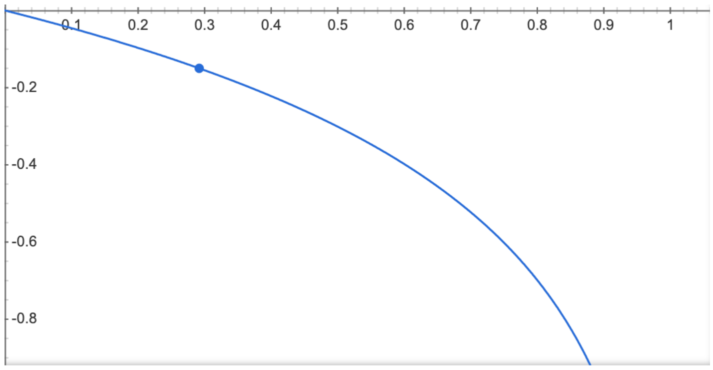

Figure 1: Plot of log(1-x)

  to $\log(1 - D(G(x))$. As $D(x)$ approaches 0, the gradient of that function grows smaller, hindering the convergence of the optimisation procedure (stochastic gradient descent).

- **b** Substituting $D(\mathbf{x})$ with ξ and $p_r(\mathbf{x})$ and $p_g(\mathbf{x})$ with a and b, the integrand is:

$$f(\xi) = a\log(\xi) + b\log(1-\xi)$$

  Setting the derivative of f with respect to ξ zero yields:

$$\frac{\partial}{\partial\xi} f = 0 \Leftrightarrow \frac{a}{\xi} - \frac{b}{1-\xi} = 0 \Rightarrow \xi^* = \frac{a}{a+b}$$

  It follows that $D^*(\mathbf{x}) = \frac{p_r(\mathbf{x})}{p_r(\mathbf{x})+p_g(\mathbf{x})}$

- **c** A divergence in the context of statistics measures the distance between two distributions. Formally, given a set of probability distributions $\mathcal{P}$ on a random variable X, a divergence is defined as a function $\mathrm{D}[\cdot||\cdot] : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ such that $\mathrm{D}[P||Q] \geq 0$ for all $P, Q \in \mathcal{P}$, and $\mathrm{D}[P||Q] = 0$ iff. (if and only if) $P = Q$.

- **d** Plugging the solution for the perfect discriminator D^* into the objective we get

$$V(D^*, G) = \int_{\mathbf{x}} p_r(\mathbf{x}) \log\left(\frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})}\right) + p_g(\mathbf{x}) \log\left(1 - \frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})}\right) d\mathbf{x}$$
$$\Leftrightarrow \int_{\mathbf{x}} p_r(\mathbf{x}) \log\left(\frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})}\right) + p_g(\mathbf{x}) \log\left(\frac{p_g(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})}\right) d\mathbf{x}$$

The JS-Divergence is

$$
\begin{aligned}
D_{JS} &= \frac{1}{2}D_{KL}(p_r \| \frac{p_r + p_g}{2}) + \frac{1}{2}D_{KL}(p_g \| \frac{p_r + p_g}{2}) \\
&\Leftrightarrow \frac{1}{2}\left(\int_{\mathbf{x}} p_r(\mathbf{x}) \log\left(\frac{2p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})} d\mathbf{x} \right) + \frac{1}{2}\left(\int_{\mathbf{x}} p_g(\mathbf{x}) \log\left(\frac{2p_g(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})} d\mathbf{x} \right)\right.\right. \\
&\Leftrightarrow \frac{1}{2}\left(\log 2 + (\int_{\mathbf{x}} p_r(\mathbf{x}) \log\left(\frac{p_r(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})} d\mathbf{x} \right) + \frac{1}{2}\left(\int_{\mathbf{x}} p_g(\mathbf{x}) \log\left(\frac{p_g(\mathbf{x})}{p_r(\mathbf{x}) + p_g(\mathbf{x})} d\mathbf{x} \right)\right.\right. \\
&\Leftrightarrow \log 4 + \frac{1}{2}V(D^*, G) \rightarrow V(D^*, G) = 2D_{JS}(p_r \| p_g) - \log 4
\end{aligned}
$$

2 Question 2

- **a**
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- **b** The MLE objective is
$$\boldsymbol{\theta}^* = \arg\max \frac{1}{N} \sum_{n=1}^{N} \log p_{\boldsymbol{\theta}}(\mathbf{x}_n).$$

  In order to compute (and differentiate) the likelihood, we must compute:
$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \int p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

  Since $p_{\boldsymbol{\theta}}(\mathbf{x}$ is presented by a neural network and therefore complex, there is no closed-form solution for this integral, making it intractable.

- **c** We have $\Sigma_1 = \boldsymbol{I}$ and $\Sigma_{\mathbf{0}} = \text{diag}(\sigma_1^2, ..., \sigma_i^2)$. Therefore,
$$\text{tr}\left(\Sigma_{\mathbf{1}}^{\mathbf{-1}}\Sigma_{\mathbf{0}}\right) = \sum_{i=1}^{K} \sigma_i^2$$
  and
$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = -\boldsymbol{\mu}_0^T \cdot (-\boldsymbol{\mu}_0) = \sum_{i=1}^{K} \mu_i^2$$
  and
$$\log\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right) = \log\left(\frac{1}{\Pi_{i=1}^K \sigma_i^2}\right) = \sum_{i=1}^{K} \log\left(\frac{1}{\sigma_i^2}\right) = -\sum_{i=1}^{K} \log(\sigma_i^2)$$