Aprendizado Profundo (AP) – 2025/2 – Prova 4 (12 pts)
Prof. Flavio Figueiredo

Nome: Felipe Bartelt de Assis Pessoa
Matrícula: 2025690830

INSTRUÇÕES
(a) As respostas devem estar na prova no espaço designado para tal.
(b) A interpretação das questões faz parte da prova. Explique as suposições que fizer.
(c) Não é permitido o uso de material de consulta.

Questão 3 (6 pts) Liste **três** diferenças principais entre um Autoencoder (AE) tradicional e um Variational Autoencoder (VAE), abordando:

- A função objetivo (loss)
- A estrutura do espaço latente
- Aplicações típicas

Inclua também uma vantagem do uso de VAEs em relação a AEs.

Resposta: • Um AE apenas minimiza a perda de reconstrução entre a entrada original e a saída reconstruída, ex: $\mathcal{L}_{AE} = \|x - g(f(x))\|^2$. No entanto, um VAE maximiza o ELBO, que combina o termo de reconstrução e um termo de regularização (divergência KL) entre a distribuição latente aprendida e uma a priori, ex: $\mathcal{L}_{VAE} = \mathbb{E}_{q_\varphi(z|x)}[\log p_\theta(x|z)] - KL(q_\varphi(z|x)\|p(z))$.

• AE tem espaço latente determinístico, enquanto VAE tem espaço latente probabilístico

• AEs são usados para compressão de dados, remoção de ruído e reconstruções. Os VAEs são mais aplicados para geração de dados, ex: imagens, voz e moléculas.

Os VAEs permitem gerar novas amostras realistas a partir de uma distribuição latente contínua e suave, enquanto os AEs só são capazes de reconstruir a entrada, não sendo capazes de gerar novas amostras realistas.

Questão 6 (6 pontos) A função de perda (loss) de um VAE é dada por:

$$\mathcal{L} = -\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z}))$$

Por que não é possível fazer backpropagation diretamente sobre amostras de $q_\phi(\mathbf{z}|\mathbf{x})$, e como esse problema é contornado?

Resposta: Ao calcular o gradiente desse funcional em relação a φ, há o problema que a distribuição $q_\varphi(z|x)$ depende de φ e a variável z é amostrada de um processo estocástico. É justamente neste ponto que a derivada não é bem definida, impedindo o back propagation.

Para contornar esse problema, usa-se uma estratégia de reparametrização de z por uma transformação diferenciável e invertível de uma outra variável aleatória ε: $z = g(\varphi, x, \varepsilon)$ tal que a distribuição de ε seja independente de x e φ. Se $q_\varphi(z|x) = \mathcal{N}(z; \mu(x;\varphi), \sigma^2(x;\varphi))$, uma reparametrização possível é $p(\varepsilon) = \mathcal{N}(\varepsilon; 0, I)$; $z = \mu(x;\varphi) + \sigma(x;\varphi) \odot \varepsilon$. Como a amostragem é desacoplada do parâmetro φ (aleatoriedade vem de ε), a esperança se torna em relação a $p(\varepsilon)$ e se torna possível calcular gradientes com relação a φ, já que z é diferenciável em φ. Por fim, o gradiente é estimado via integração de Monte Carlo.