1 Variational autoencoders

Suppose the following latent variable model  

\[ p(x) = \int_z p(z)p(x|z). \]

The entropy of a density \( q \) is defined as  

\[ H(q) = -\int_z q(z) \ln q(z). \]

The Kullback Leibler divergence between densities \( p \) and \( q \) is defined as  

\[ D(q||p) = \int_z q(z) \ln \frac{q(z)}{p(z)}. \]

(a) Show that  

\[ F = \int z q(z|x) \ln p(x|z) - D(q(z|x)||p(z)) \leq \ln p(x) \]

(b) Suppose the generative model \( p(x) \) is fixed, write the objective function \( F \) in the form of a single KL divergence that involves the encoder \( q(z|x) \), and a second term that does not depend on the encoder.

(c) What are the two reasons that the variational bound from above is in general not tight?

(d) Describe the re-parametrization trick, and why it is used in variational autoencoders.

2 Other generative models

(a) The objective function of a Generative Adversarial Network is given by  

\[ V(\phi, \theta) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\ln D(x)] + \mathbb{E}_{z \sim p(z)}[\ln (1 - D(\phi(G(\theta(z))))]. \]

Derive the optimal discriminator \( D(x) \) that maximizes the objective function, writing it in terms of \( p_{\text{data}}(x) \) and \( p_{\text{GAN}}(x) = \int_z p(z) \delta(x - G(\theta(z))) \).

(b) Describe the invertible layers used in the “Non-volume preserving” (NVP) model. How is the inverse of the layer computed?
(c) Explain how the Jacobian determinant of such a layer in NVP is computed, and what it’s time complexity is.

(d) Describe one similarity between a classic autoencoder and the “Masked Autoencoder for Distribution Estimation” (MADE) model. Describe also two differences.

3 Are you positive definite?

Tell if the kernels below are positive definite or not. For each example, provide a simple proof.

\[
\begin{align*}
&\forall x, y \geq 0 \quad K_1(x, y) = \min(x, y) \\
&\forall x, y \geq 0 \quad K_2(x, y) = \max(x, y) \\
&\forall x, y > 0 \quad K_3(x, y) = \frac{\min(x, y)}{\max(x, y)} \\
&\forall x, y > 0 \quad K_4(x, y) = \frac{\max(x, y)}{\min(x, y)} \\
&\forall x, y, \alpha > 0 \quad K_5(x, y) = e^{-\alpha(x-y)^2} \\
&\forall x, y \in (-1, 1) \text{ (interval excluding -1 and 1)} \quad K_6(x, y) = \frac{1}{1-xy} \\
&\forall x, y \in \mathbb{R} \quad K_7(x, y) = \cos(x-y) \\
&\forall x, y \in \mathbb{R} \quad K_8(x, y) = \cos(x+y) \\
&\forall x, y \in \mathbb{R} \quad K_9(x, y) = 10^{xy} \\
&\forall x, y \in \mathbb{R} \quad K_9(x, y) = 10^{x+y}
\end{align*}
\]

4 Combination rules of kernels

Consider two positive definite kernels \( K_1 \) and \( K_2 \).

(a) Show that for all \( \alpha, \beta > 0 \), \( \alpha K_1 + \beta K_2 \) is positive definite

(b) Show that the element wise product \( K_1(x, y)K_2(x, y) \) is positive definite.

5 What is my RKHS?

(a) Recall the definition of an RKHS.

(b) What is the RKHS of the polynomial kernel of degree 2: \( K(x, y) = (x^\top y)^2 \) for \( x, y \in \mathbb{R}^d \)? Give a proof.

(c) What is the RKHS of the polynomial kernel of degree 3: \( K(x, y) = (xy)^3 \) for \( x, y \in \mathbb{R} \)? Give a proof.

6 Kernel for equivalence classes

Consider a similarity measure \( K : \times \to \{0, 1\} \) with \( K(x, x) = 1 \) for all \( x \) in . Prove that \( K \) is p.d. if and only if, for all \( x, x', x'' \) in ,

- \( K(x, x') = 1 \iff K(x', x) = 1 \), and
- \( K(x, x') = K(x', x'') = 1 \Rightarrow K(x, x'') = 1 \).
7 A more difficult exercise

1. Let
\[ H = f : [0, 1] \to \text{absolutely continuous}, f' \in L^2([0, 1]), f(0) = 0, \]
endowed with the bilinear form
\[ \forall f, g \in H, \quad (f, g)_{H} = \int_{0}^{1} f'(u)g'(u)du. \]
Show that \( H \) is an RKHS, and compute its reproducing kernel.

2. same question when
\[ H = f : [0, 1] \to \text{absolutely continuous}, f' \in L^2([0, 1]), f(0) = f(1) = 0, \]

3. Same question, when \( H \) is endowed with the bilinear form:
\[ \forall f, g \in H, \quad (f, g)_{H} = \int_{0}^{1} f(u)g(u) + f'(u)g'(u)du. \]

Absolute continuity is a notion that is stronger than continuity. \( f \) is said to be absolutely continuous on \([0, 1]\) if and only if \( f \) has a derivative \( f' \) almost everywhere on \([0, 1]\), \( f' \) is Lebesgue integrable and for all \( x, y > 0 \),
\[ f(x) = f(0) + \int_{t=0}^{x} f'(t)dt. \]