1. (a) Complete the proof of Theorem 4.7 by showing that if \( k_1 \) and \( k_2 \) are valid kernels, then so are \( k_1 + k_2 \) and \( ak_1 \) for \( a > 0 \).

(b) Assume that \( k \) is a valid kernel with feature map \( \psi \). Obtain \( \tilde{k} \) by normalising:

\[
\tilde{k}(x, z) = \frac{k(x, z)}{(k(x, x)k(z, z))^{1/2}}.
\]

Show that \( \tilde{k} \) is a kernel for the feature map given by \( \tilde{\psi}(x) = \psi(x)/\|\psi(x)\| \).


Hints: For part 1, use induction. For part 2, notice that \( \exp(r) \) is a limit of polynomials \( p_n(r) \).

For part 3, normalise the kernel \( \exp(\langle x, z \rangle/\sigma^2) \).

3. Define \( f(x) = x^2 \) and \( g(x) = (x-2)^2 \). Consider the problem

\[
\text{minimise } f(x) \text{ subject to } g(x) \leq 1.
\]

Write out the Lagrangian \( L(x, \lambda) \). Solve the problem using the KKT conditions. Write out also the dual \( g(\lambda) \) in a closed form.

Use some mathematical software (e.g., Maple) to plot \( L(x, \lambda) \) (as a function of two variables, so you get a three-dimensional picture). Find from this plot the functions \( x \mapsto \max_\lambda L(x, \lambda) \) and \( \lambda \mapsto \min_x L(x, \lambda) \). Notice the point where they coincide.

Repeat the exercise with the constraint \( g(x) \leq 8 \).

4. In an online prediction setting, given an example \((x_t, y_t) \in \mathbb{R}^n \times \mathbb{R}\), let \( W_t = \{ w \in \mathbb{R}^n \mid (w \cdot x_t - y_t)^2 \leq R^2 \} \) where \( R \) is some constant. Then define the update as

\[
w_{t+1} = \arg \min_{w \in W_t} \|w_t - w\|^2.
\]

Write out the update in a closed form (i.e., solve the minimisation).

5. Recall the relative entropy

\[
d_{re}(p, q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}.
\]

where we assume \( p_i, q_i \geq 0 \) and \( \sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1 \). Fix \( q \) such that \( q_i > 0 \) for all \( i \). Using the KKT conditions, show that subject to the constraints \( p_i \geq 0, i = 1, \ldots, n \), and \( \sum_{i=1}^n p_i = 1 \), the relative entropy \( d_{re}(p, q) \) is minimised at \( p = q \). You may take it as known that \( d_{re}(p, q) \) is convex in \( p \) (and also \( q \), although that is not needed here.)

What is the dual?