1. [40 points] The entropy of a distribution \( p(x) \) is given by

\[
H[x] = -\int p(x) \log p(x) \, dx
\]

Derive the entropy of the univariate Gaussian \( \mathcal{N}(x|\mu, \sigma^2) \):

\[
p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]
\]

2. [20 points] In this problem, you will directly show that the multivariate distribution with maximum entropy, for a given mean and covariance, is a Gaussian. We wish to maximize \( H[x] \) over all distributions subject to the constraints that the distribution is normalized and has a specific mean and covariance, i.e.

\[
\int p(x) \, dx = 1
\]

\[
\int xp(x) \, dx = \mu
\]

\[
\int (x - \mu)(x - \mu)^\top p(x) \, dx = \Sigma
\]

Show that solving this constrained optimization problem yields \( p(x) = \mathcal{N}(x|\mu, \Sigma) \) (Hint: use Lagrange multipliers. Do not use the derivation steps covered in the lecture.)

3. [20 points] Consider a linear model of the form

\[
y(x, w) = w_0 + \sum_{i=1}^{D} w_i x_i
\]

together with a sum-of-squares error function of the form

\[
error_D(w) = \frac{1}{2} \sum_{t=1}^{N} (y(x_t, w) - r_t)^2
\]

Now suppose that Gaussian noise \( \epsilon_t \) with zero mean and variance \( \sigma^2 \) is added independently to each of the input variables \( x_i \). Show that minimizing \( error_D \) averaged over the noise distribution is equivalent to minimizing the sum-of-squares error for noise-free input variables with the addition of some weight-decay regularization term.

4. [20 points] Derive the formula for basis function \( \phi(x) \) that is equivalent to the quadratic kernel

\[
k(x, x') = (1 + x^\top x')^2
\]