Homework 3

Aukkawut Ammartayakun
CS 539 Machine Learning

Spring 2023

Problem 1

Evaluate the Kullback-Leibler divergence (equation 1.113 in Bishop)

$$\mathrm{KL}(p||q) = -\int p(x) \ln \left[\frac{p(x)}{q(x)}\right] \, dx$$

between two Gaussians $p(x) = \mathcal{N}(x|\mu, \Sigma)$ and $q(x) = \mathcal{N}(x|m, L)$.

Solution. Assume that p and q are n dimensional gaussian;

$$\begin{aligned}\int p(x) \ln \left[\frac{p(x)}{q(x)}\right] \, dx &= \frac{1}{2}\left(\ln \frac{|\Sigma|}{|L|} - \mathbb{E}[(x-\mu)^\top \Sigma^{-1}(x-\mu)] + \mathbb{E}[(x-m)^\top L^{-1}(x-m)]\right) \\ &= \frac{1}{2}\left[\ln\left(\frac{|\Sigma|}{|L|} - n + (\mu-m)^\top \Sigma^{-1}(\mu-m) + \mathrm{tr}(L^{-1}\Sigma)\right)\right]\end{aligned}$$

Problem 2

Suppose that $p(x)$ is some fixed distribution and that we wish to approximate it using a Gaussian distribution $q(x) = \mathcal{N}(x|\mu, \Sigma)$. By writing down the form of the KL divergence $\mathrm{KL}(p||q)$ for a Gaussian $q(x)$ and then differentiating, show that minimization of $\mathrm{KL}(p||q)$ with respect to μ and Σ leads to the result that μ is given by the expectation of x under $p(x)$ and that Σ is given by the covariance.

Solution. Given that $q(x)$ is gaussian,

$$\mathrm{KL}(p||q) = \mathbb{E}_p[\ln q(x)] + k$$

for some constant k that represent negative entropy of $p(x)$. Differentiate this and we will get that μ is given by the expectation of x under $p(x)$ and that Σ is given by the covariance by default.

Problem 3

Consider a regression problem involving multiple target variables in which it is assumed that the distribution of the targets, conditioned on the input vector x, is a Gaussian of the form

$$p(t|x, w) = \mathcal{N}(t|y(x, w), \Sigma)$$

where $y(x, w)$ is the output of a neural network with input vector x and weight vector w, and Σ is the covariance of the assumed Gaussian noise on the targets.

Given a set of independent observations of x and t, write down the error function that must be minimized in order to find the maximum likelihood solution for w, if we assume that Σ is fixed and known. Now assume that Σ is also to be determined from the data, and write down an expression for the maximum likelihood solution for Σ. Note that the optimizations of w and Σ are now coupled, in contrast to the case of independent target variables discussed in Section 5.2.

Solution. Assume that y and t is k dimensional vector, log-likelihood in this case is

$$\ln p(t|x, w) = -\frac{N}{2}\ln|\Sigma| - \frac{1}{2}(t - y(x, w))^\top \Sigma^{-1}(t - y(x, w)) - \frac{Nk}{2}\ln(2\pi)$$

Using the profiling likelihood, fixed w. Evaluate the derivative with respect to Σ and set it to zero will yield use

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(t_i - y_i(x_i, w))(t_i - y_i(x_i, w))^\top$$

Problem 4

Let $z = e^{x^2 y}$ where $x(u; v) = \sqrt{uv}$ and $y(u; v) = \frac{1}{v}$.

1. Derive $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$

   Solution.

$$\begin{aligned}
\frac{\partial z}{\partial u} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} \\
&= 2xye^{x^2 y}\frac{v}{2\sqrt{uv}} \\
\frac{\partial z}{\partial v} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v} \\
&= 2xye^{x^2 y}\frac{v}{2\sqrt{uv}} + x^2 e^{x^2 y}\frac{-1}{v^2}
\end{aligned}$$

2. Let's assume the target value for the output (z) is t. We want to minimize the $e = \frac{1}{2}(t - z)^2$; write down the update rule for changing u and v that minimizes e.

   Solution. The graph is as followed $u, v \to x, y \to z \to e$ Evaluate $\frac{\partial e}{\partial u} = -(t-z)\frac{\partial z}{\partial u}$ and $\frac{\partial e}{\partial v} = -(t-z)\frac{\partial z}{\partial v}$. As shown in the first part, we can use gradient descent to update u and v with the gradient from error calculated aboved. That is

$$\begin{aligned}
u_{n+1} &= u_n - \alpha\frac{\partial e}{\partial u} \\
v_{n+1} &= v_n - \alpha\frac{\partial e}{\partial v}
\end{aligned}$$

   for constant α.