Machine Learning: Homework #12

Due on January 29, 2018 at 09:59am

Professor Dr. Stephan Guennemann

Marc Meissner - 03691673

Problem 1

Compute the KL divergence between two Gaussian distributions $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ with diagonal covariance matrices.

Solution

For D-dimensional diagonal gaussians (no coupling), the pdf decomposes into a product of the individual components/dimensions:

$$p(z) = \prod_j p_j(z_j) = \prod_j N(z_j|\mu_{1,j}, \sigma_{1,j}^2)$$

This helps us in decomposing the problem into D independent sums or subproblems:

$$\mathrm{KL}(p||q) = \int_z p(z) \log \frac{p(z)}{q(z)} dz = \mathrm{E}_{z \ p(z)}[\log(\prod_j N(z_j|\mu_{1,j}, \sigma_{1,j}^2) - \log(\prod_j N(z_j|\mu_{2,j}, \sigma_{2,j}^2))]$$

$$= \sum_j \mathrm{E}_{z \ p(z)}[\log(N(z_j|\mu_{1,j}, \sigma_{1,j}^2) - \log(N(z_j|\mu_{2,j}, \sigma_{2,j}^2))] = \sum_j \mathrm{KL}(p_j||q_j)$$

For these individual subproblems, the solution is:

$$\mathrm{KL}(p||q) = \sum_j \mathrm{KL}(p_j||q_j) = -\frac{n}{2} + \sum_j \left(\ln \frac{\sigma_{2,j}}{\sigma_{1,j}} + \frac{\sigma_{1,j}^2 + (\mu_{1,j} - \mu_{2,j})^2}{2\sigma_{2,j}^2} \right)$$

Problem 2

Consider that $p(x)$ is some arbitrary fixed distribution that we wish to approximate using an isotropic Gaussian distribution $q(x) = N(x|\mu, I)$ (covariance matrix is identity matrix). By writing down the KL divergence $\mathrm{KL}(p||q)$ and then differentiating w.r.t. μ, show that the optimal setting of the parameter is $\mu^* = \mathrm{argmin}\mathrm{KL}(p||q) = \mathrm{E}_p[x]$.

Solution

$$\mathrm{KL}(p||q) = \int_x p(x)(\log p(x) - \log q(x))dx$$

$$\frac{d}{d\mu}\mathrm{KL}(p||q)$$

$$= \frac{d}{d\mu}\left(-\int_x p(x)\log q(x)\right)dx$$

$$= \frac{d}{d\mu}\left(-\int_x p(x)\frac{k}{2}\log(2\pi)dx + \int_x p(x)\frac{1}{2}(x-\mu)^T(x-\mu)dx\right)$$

$$= \frac{1}{2}\int_x \frac{d}{d\mu} p(x)(x-\mu)^T(x-\mu)dx$$
$$= \frac{1}{2}\int_x p(x)(-2(x-\mu))dx$$
$$= -\int_x p(x)xdx + \int_x p(x)\mu dx \overset{!}{=} 0$$

From this, it follows:

$$\mathrm{E}[x] = \mu$$

Problem 3

Consider a very simple probabilistic model with a 2-D latent variable $z \in R^2$ and an observed variable $x \in R$. The prior over the latent variable is $p(z) = N(z|0, I) = N(z_1|0, 1) \cdot N(z_2|0, 1)$, and the likelihood is $p(x|z) = N(x|\theta^T z, 1)$, where $\theta \in R^2$ is a known and fixed parameter.

Write down the true posterior distribution $p(z|x)$. Can the posterior be factorized over z_1 and z_2 ? (i.e. can it be expressed as $p(z_1|x)p(z_2|x)$?)

Solution

$$p(z|x) \propto p(x|z)p(z)$$
$$= \frac{1}{(\sqrt{2\pi})^3}\exp\left(-\frac{1}{2}\left((x-\theta_1 z_1 - \theta_2 z_2)^2 + z_1^2 + z_2^2\right)\right)$$

One can immediately see that the coupling between z_1 and z_2 makes it impossible to factorize the posterior.

Problem 4

We approximate the true posterior using a mean-field variational distribution $q(z) = q_1(z_1)q_2(z_2) = N(z_1|m_1, s_1^2)\cdot N(z_2|m_2, s_2^2)$.
Your task is to derive the optimal updates for q_1 and q_2 . Is $q(z)$ able to match the true posterior $p(z|x)$?

Solution

Coefficient matching yields that a factorizable pdf $q(z)$ can not be equal to a posterior $p(z|x)$ which is generally not factorizable. Thus, there will be an approximation error.
Regardless, one can at least do his best, right?

Starting with the ELBO:

Problem 4 continued on next page...

$$L(q) = \mathrm{E}_q[\log(p(z,x)) - \log(q(z))]$$
$$= \mathrm{E}_q[\log(p(z))] + \mathrm{E}_q[\log(p(x|z))] - \mathrm{E}_q[\log(q_1(z_1))] - \mathrm{E}_q[\log(q_2(z_2))]$$
$$= -\frac{1}{2}[(m_1^2+s_1^2)(1+\theta_1^2)+(m_2^2+s_2^2)(1+\theta_2^2)+\log(2\pi)+x^2]-\theta_1\theta_2 m_1 m_2+x\theta_1 m_1+x\theta_2 m_2+\log(s_1)+\log(s_2)$$

Computing the derivative w.r.t. m_1 and s_1, respectively:

$$\frac{\partial L(q)}{\partial m_1} = -m_1(1+\theta_1^2) - \theta_1\theta_2 m_2 + x\theta_1 = 0$$
$$m_1 = \theta_1 \frac{x - \theta_2 m_2}{1+\theta_1^2}$$
$$\frac{\partial L(q)}{\partial s_1} = -s_1(1+\theta_1^2) + \frac{1}{s_1} = 0$$
$$s_1^2 = \frac{1}{1+\theta_1^2}$$

m_2 and s_2^2 are derived in the same way.