Machine Learning Assignment 12

Shivangi Aneja

29-January-2018

Problem 1

KL divergence for 2 Gaussians :

$$\mathcal{N}(x|\mu_1,\Sigma_1)\ ,\ \mathcal{N}(x|\mu_2,\Sigma_2)$$

$KL(p||q) = \int p(x)\ log\dfrac{p(x)}{q(x)}dx = \int p(x)[log(p(x)) - log(q(x))]dx$
$\Rightarrow \int p(x)\Big[\Big(\dfrac{D}{2}log\dfrac{1}{2\pi} + \dfrac{1}{2}log\dfrac{1}{|\Sigma_1|} - \dfrac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\Big) + \Big(-\dfrac{D}{2}log\dfrac{1}{2\pi} - \dfrac{1}{2}log\dfrac{1}{|\Sigma_2|} + \dfrac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\Big)$
$\Rightarrow \dfrac{1}{2}\int p(x)\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - (x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + (x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\Big]dx$
$\Rightarrow \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|}\int p(x)dx - \int(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)p(x)dx + \int(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)p(x)dx\Big]$
$\Rightarrow \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - E_{p(x)}[(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)] + E_{p(x)}[(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)]\Big]$

Using Eq(380) of Matrix Cookbook :

$$E[(x-m)^TA(x-m)] = (\mu-m)^TA(\mu-m) + Tr(A\Sigma)$$

$\Rightarrow \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - [Tr(\Sigma_1^{-1}\Sigma_1) + (\mu_1-\mu_1)^T\Sigma_1^{-1}(\mu_1-\mu_1)] + [Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)]\Big]$
$\Rightarrow \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - [Tr(I_d) + (0)] + [Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)]\Big]$
$\Rightarrow \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - D + Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)\Big]$

$$\boxed{KL(p||q) = \dfrac{1}{2}\Big[log\dfrac{|\Sigma_2|}{|\Sigma_1|} - D + Tr(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)^T\Sigma_2^{-1}(\mu_2-\mu_1)\Big]}$$

Problem 2

$p(x) \approx q(x) = \mathcal{N}(x|\mu, I)$

$KL(p||q) = \int p(x)\ log\dfrac{p(x)}{q(x)}dx = \int p(x)[log(p(x)) - log(q(x))]dx$

$\Rightarrow \int p(x)log(p(x))dx - \int p(x)log(q(x))dx$

$\Rightarrow \int p(x)log(p(x))dx - \int p(x)log\Big[\dfrac{1}{(2\pi)^{D/2}|I|}exp(-(x-\mu)^T I^{-1}(x-\mu))\Big]dx$

Using $|I| = 1$ and $I^{-1} = I$

$\Rightarrow \int p(x)log(p(x))dx - \dfrac{D}{2}log\dfrac{1}{2\pi} + \int (x-\mu)^T I(x-\mu)p(x)dx$

$\Rightarrow \boxed{KL(p||q) = \int p(x)log(p(x))dx - \dfrac{D}{2}log\dfrac{1}{2\pi} + \int (x^T x - 2x^T\mu + \mu^T\mu)p(x)dx}$

To find the optimal parameter of μ we have,

$$\mu* = argmin_{\mu} KL(p||q)$$

Now setting derivative of $KL(p||q)$ to zero,

$\dfrac{\partial KL(p||q)}{\partial \mu} \doteq 0$

$\Rightarrow \int(-2x^T + 2\mu)p(x)dx = 0$

$\Rightarrow -E_{x\sim p(x)}[x] + \mu E_{x\sim p(x)}[1] = 0 \Rightarrow -E_{x\sim p(x)}[x] + \mu = 0$

$\Rightarrow \boxed{\mu* = E_{p(x)}[x]}$

Thus,

$\boxed{\mu* = argmin_{\mu} KL(p||q) = E_{p(x)}[x]}$

Problem 3
Given 2-D latent variables $z \in \mathcal{R}^2$
Observed Variable $x \in \mathcal{R}$
$\theta \in \mathcal{R}^2$
The prior over latent is :

$$p(z) = \mathcal{N}(z|0, I) = \mathcal{N}(z_1|0, 1)\mathcal{N}(z_2|0, 1)$$

The likelihood is :

$$p(x|z) = \mathcal{N}(x|\theta^T z, 1)$$

The posterior is given by :
$p(z|x) = \dfrac{p(x|z)p(z)}{p(x)} \propto \mathcal{N}(x|\theta^T z, 1) \times \mathcal{N}(z|0, I)$
$p(z|x) \propto \mathcal{N}(x|\theta^T z, 1) \times \mathcal{N}(z_1|0, 1) \times \mathcal{N}(z_2|0, 1)$
Upto a normalizing constant , we have :
$p(z|x) = exp\Big(-\dfrac{(x-\theta^T z)^2}{2}\Big)exp\Big(-\dfrac{z_1^2}{2}\Big)exp\Big(-\dfrac{z_2^2}{2}\Big)$
$= exp\Big(-\dfrac{x^2 + (\theta^T z)^2 - 2(\theta^T z)x + z_1^2 + z_2^2}{2}\Big)$
$= exp\Big(-\dfrac{x^2 + (\theta_1 z_1 + \theta_2 z_2)^2 - 2x(\theta_1 z_1 + \theta_2 z_2) + z_1^2 + z_2^2}{2}\Big)$
$= exp\Big(-\dfrac{x^2 + \theta_1^2 z_1^2 + \theta_2^2 z_2^2 + 2\theta_1\theta_2 z_1 z_2 - 2x(\theta_1 z_1 + \theta_2 z_2) + z_1^2 + z_2^2}{2}\Big)$

Thus, the posterior can not be factorized over z_1 and z_2 as the exponent contains $2\theta_1\theta_2 z_1 z_2$.

Problem 4

Consider a Gaussian distribution

$$p(z) = \mathcal{N}(z|\mu, \lambda^{-1})$$

over two correlated variables $z = (z_1, z_2)$ in which the mean and precision have elements

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_1 \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

$$\begin{aligned} \ln q^*_{z_1}(z_1) &= \mathrm{E}_{z_2}\left[\ln p(z)\right] + C \\ &= \mathrm{E}_{z_2}\left[-\frac{(z_1-\mu_1)^2\lambda_{11}}{2} - (z_1-\mu_1)\lambda_{12}(z_2-\mu_2)\right] + C \end{aligned}$$

Solving it further we get

$$= -\frac{z_1^2\lambda_{11}}{2} + z_1\mu_1\lambda_{11} - z_1\lambda_{12}(E[z_2] - \mu_2) + C$$

We observe that the right-hand side of this expression is a quadratic function of z_1, and so we can identify $q(z_1)$ as a Gaussian distribution.

We can write this as:

$$\begin{aligned} q^*_{z_1}(z_1) &\sim \mathcal{N}(z_1 \mid m_1, \lambda_{11}^{-1}) \\ m_1 &= \mu_1 - \lambda_{11}^{-1}\lambda_{12}(E[z_2] - \mu_2 \end{aligned}$$

By symmetry

$$\begin{aligned} q^*_{z_2}(z_2) &\sim \mathcal{N}(z_2 \mid m_2, \lambda_{22}^{-1}) \\ m_2 &= \mu_2 - \lambda_{22}^{-1}\lambda_{21}(E[z_1] - \mu_1 \end{aligned}$$

But these Gaussians are uncorrelated, so off-diagonal entries become zero. So, $\lambda_{12} = \lambda_{21} = 0$

So, we have

$$\begin{aligned} q^*_{z_1}(z_1) &\sim \mathcal{N}(z_1 \mid m_1, \lambda_{11}^{-1}) \\ m_1 &= \mu_1 \end{aligned}$$

By symmetry

$$\begin{aligned} q^*_{z_2}(z_2) &\sim \mathcal{N}(z_2 \mid m_2, \lambda_{22}^{-1}) \\ m_2 &= \mu_2 \end{aligned}$$