Machine Learning

Homework 12 solution - Shyam Arumugaswamy

1.

Let f(x) and g(x) denote the respective densities. Assuming the notation

$$\mu_{\mathrm{i}} = (\mu_{\mathrm{i},1}, \ldots, \mu_{\mathrm{i},\mathrm{n}}), \quad \Sigma_{\mathrm{i}} = \mathrm{diag}({\sigma^2}_{\mathrm{i},1}, \ldots., {\sigma^2}_{\mathrm{i},\mathrm{n}})$$

For diagonal Gaussian, the components are independent and the pdf decomposes:

$$\mathrm{f(x)} = \prod_{\mathrm{j}} \mathrm{f}_{\mathrm{j}}\, (\mathrm{x}_{\mathrm{j}}) = \prod_{\mathrm{j}} \mathrm{N}\, (\mathrm{x}_{\mathrm{j}} |\, \mu_{1,\mathrm{j}},\ {\sigma^2}_{\mathrm{i},\mathrm{j}})$$

and similarly for g.

$$\mathrm{KL(f\ ||\ g)} = \int \mathrm{f(x)}\, \ln(\frac{\mathrm{f(x)}}{\mathrm{g(x)}})\, \mathrm{dx}$$

Since f and g factorize , the logarithm of the fraction turns into a sum of fractions. Linearity of expectation then gives us that the KL decomposes into a sum of KL divergences of the components:

$$\mathrm{KL(f\ ||\ g)} = \sum_{\mathrm{j}} \mathrm{KL}(\mathrm{f}_{\mathrm{j}}\, ||\, \mathrm{g}_{\mathrm{j}})$$

We have reduced this to one dimensional case

$$\mathrm{KL(f\ ||\ g)} = \sum_{\mathrm{j}} \mathrm{KL}(\mathrm{f}_{\mathrm{j}}\, ||\, \mathrm{g}_{\mathrm{j}}) = \frac{-\mathrm{n}}{2} + \sum_{\mathrm{j}}(\ln \frac{\sigma_{2,\mathrm{j}}}{\sigma_{1,\mathrm{j}}} + \frac{{\sigma^2}_{1,\mathrm{j}} + (\mu_{1,\mathrm{j}} - \mu_{2,\mathrm{j}})^2}{2{\sigma_{2,\mathrm{j}}}^2})$$