Question: 2. Consider that p= N(H. 2) and q = N(0,1). Here I = diag(01.02....

2. Consider that $p = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $q = \mathcal{N}(\mathbf{0}, \mathbf{I})$. Here $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$. Then, shown that the KL divergence between p and q is defined as below.

$$D_{\mathrm{KL}}(p|q) = \frac{1}{2}\sum_{i=1}^{k}\left(\sigma_i^2 + \mu_i^2 - 1 - \log_e(\sigma_i^2)\right).$$

Show transcribed image text

Expert Answer

Anonymous answered this
58 answers

Was this answer helpful? 1 0

Ans.

KL divergence between two distribution P & Q of a continuous random variable is given by.

$$D_{KL}(P\|Q) = \int_x P(x) \log \frac{P(x)}{q(x)}$$

$P = N(\mu, \Sigma)$ and $q = N(0, I)$

$$P(x) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$

$$D_{KL}(P\|Q) = E_p[\log(P) - \log(Q)]$$

$$= E_p\left[\frac{1}{2}\log\frac{|\Sigma_q|}{|\Sigma_p|} - \frac{1}{2}(x-\mu_p)^T \Sigma_p^{-1}(x-\mu_p) + \frac{1}{2}(x-\mu_q)^T \Sigma_q^{-1}(x-\mu_q)\right]$$

$$= \frac{1}{2}\log\frac{|\Sigma_q|}{|\Sigma_p|} - \frac{1}{2}E_p\left[(x-\mu_p)^T \Sigma_p^{-1}(x-\mu_p)\right] + \frac{1}{2}E_p\left[(x-\mu_q)^T \Sigma_q^{-1}(x-\mu_q)\right].$$

Now,

$$(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1} (\underline{x}-\underline{\mu}_p) = tr\left\{(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1} (\underline{x}-\underline{\mu}_p)\right\}$$

$$= tr\left\{(\underline{x}-\underline{\mu}_p)(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1}\right\}$$

Also

$$\frac{1}{2} E_p \left[tr\{(\underline{x}-\underline{\mu}_p)(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1}\} \right].$$

$$= \frac{1}{2} tr\left\{ E_p\left[(\underline{x}-\underline{\mu}_p)(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1}\right]\right\}$$

$$= \frac{1}{2} tr\left\{ E_p\left[(\underline{x}-\underline{\mu}_p)(\underline{x}-\underline{\mu}_p)^T \Sigma_p^{-1}\right]\right\}.$$

We know,

$$E_p\left[(\underline{x}-\underline{\mu}_p)(\underline{x}-\underline{\mu}_p)^T\right] = \Sigma_p$$

$$= \frac{1}{2} tr\left\{\Sigma_p \Sigma_p^{-1}\right\}$$

$$= \frac{1}{2} tr\left\{I_k\right\}$$

$$= k/2.$$

Similarly, on simplifying.

$$E_p\left[(x-\mu_p)^T \Sigma_q^{-1}(x-\mu_q)\right] = (\mu_p-\mu_q)^T \Sigma_q^{-1}(\mu_p-\mu_q) + tr\{\Sigma_q^{-1}\Sigma_p\}.$$

Now combining all we get.

$$D_{KL}(p\|q) = \frac{1}{2}\left[\log\frac{|\Sigma_q|}{|\Sigma_p|} - k + (\mu_p-\mu_q)^T \Sigma_q^{-1}(\mu_p-\mu_q) + tr\{\Sigma_q^{-1}\Sigma_p\}.\right]$$

When q is $N(0, I)$ we get.

$$D_{KL}(p\|q) = \frac{1}{2}\left[\mu_p^T\mu_p + tr|\Sigma_p| - k - \log_e|\Sigma_p|\right]$$

~~Ak~~ We have $\Sigma = dig(\sigma_1, \sigma_2, \dots \sigma_k)$.

$$\mu_p^T\mu_p = \sum_{i=1}^{k} \mu_i^2$$

$$\log_e|\Sigma_p| = \sum_{i=1}^{k} \log_e(\sigma_i^2)$$

Putting these value in expression of D_{KL} we get:

$$D_{KL}(p\|q) = \frac{1}{2}\sum_{i=1}^{k}\left(\sigma_i^2 + \mu_i^2 - 1 - \log_e(\sigma_i^2)\right).$$

Proved.

Comment

A: See answer 92% (13 ratings)

Show more