Problem Set 5

Due Wednesday, Nov 14, 2018

Distance measure between two distributions

In the class, we introduced a distance measure between two probability density distributions, called relative entropy or Kullback-Leibler (KL) divergence,

$$D_{KL}(P\|Q) = \sum_x P(x) \log_2 \frac{P(x)}{Q(x)} \tag{1}$$

Prove that the relative entropy satisfies

$$D_{KL}(P\|Q) \geq 0 \tag{2}$$

Hint: As you have learned in calculus, a function $f(x)$ is convex if for all $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{3}$$

$\log(1/z)$ is for example a strictly convex function for $x > 0$. Prove that the above inequality can be generalized to **Jensen's inequality**:

$$\langle f(x) \rangle \geq f(\langle x \rangle), \tag{4}$$

where x is a random variable drawn from a probability distribution $P(x)$, and $\langle .. \rangle$ denotes an average, so that $\langle f(x) \rangle = \sum_x P(x) f(x)$.

Mutual Information

(a) Consider a toy model

$$y = wx + \eta, \tag{5}$$

where x and η are two gaussian random variable with zero mean and variance σ_x^2, σ_η^2 respectively, η may be viewed as the noise. Compute the mutual information between x and y and prove that it has the following expression

$$I(x, y) = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_x^2}{\sigma_\eta^2}\right) \tag{6}$$

*(b) Now consider the problem we discussed in the class

$$y_i = \sum_j W_{ij} x_j + \eta_i \tag{7}$$

Here x are random variables $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ with zero means drawn from a multivariate gaussian distribution. η_i is noise drawn also from a gaussian distribution, whose covariance matrix is given by $\langle \eta_i \eta_j \rangle = \sigma^2 \delta_{ij}$. Show that in this case, the mutual information between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$I(\mathbf{y}, \mathbf{x}) = \frac{1}{2} \mathrm{Tr} \log_2 \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{W C W^T} \right) \tag{8}$$

Note that if $\mathbf{A}$ can be diagonalized,

$$\mathbf{A} = \mathbf{U} \begin{bmatrix} \diagdown & & \\ & \lambda_i & \\ & & \diagdown \end{bmatrix} \mathbf{U}^{-1},$$

the matrix function $\log_2 \mathbf{A}$ can be defined as

$$\log_2 \mathbf{A} = \mathbf{U} \begin{bmatrix} \diagdown & & \\ & \log_2 \lambda_i & \\ & & \diagdown \end{bmatrix} \mathbf{U}^{-1}$$

Note: Problems with * are optional. However, solving them will give you additional credits.

Multivariate Gaussian distribution

Below are some mathematical notes. Consider N random variables $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ with zero means drawn from a multivariate gaussian distribution

$$p(\mathbf{x}) = \frac{1}{Z} \exp \left(-\frac{1}{2} \mathbf{x}^T \mathbf{C^{-1}} \mathbf{x} \right) \tag{9}$$

where Z is the normalization factor. $\mathbf{C}$ is the covariance matrix of $\mathbf{x}$, satisfying $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^T \rangle$. Here and below, we shall assume that $\mathbf{C}$ is positive definite, which means all its eigenvalues are real and positive. Because the covariance matrix is symmetric, all its eigenvectors $\mathbf{u_i}$ define an orthogonal basis, satisfying

$$\mathbf{u_i^T} \mathbf{u_j} = \delta_{ij} \tag{10}$$

And the covariance matrix can be written as

$$\mathbf{C}^{-1} = \sum_i \frac{1}{\lambda_i} \mathbf{u_i} \mathbf{u_i^T} \tag{11}$$

Now we shall define

$$y_i = \mathbf{u_i^T x}, \tag{12}$$

and

$$\mathbf{x}^T \mathbf{C^{-1} x} = \sum_i \frac{1}{\lambda_i} y_i^2 \tag{13}$$

We are now interested in calculating the normalization factor Z, and to do this, we need to compute the integral

$$Z = \int d\mathbf{x} \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{C^{-1} x}\right)$$

To do this, we will perform a change of coordinate from $\mathbf{x} \to \mathbf{y}$. Let's define the matrix $\mathbf{U} = [\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_N}]$, which is a unitary matrix satisfying $\mathbf{U^T U} = \mathbf{U U^T} = I$. We also have

$$\mathbf{y} = \mathbf{U^T x} \tag{14}$$

$$Z = \int |\mathbf{U}| d\mathbf{y} \exp\left(-\frac{1}{2}\sum_i \frac{1}{\lambda_i} y_i^2\right)$$

The determinant of a unitary matrix $|\mathbf{U}|$ is simply one, and all the terms within the integral are decoupled. As a result, we have

$$Z = \frac{1}{\sqrt{(2\pi)^N \prod_{i=1}^N \lambda_i}} = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{C}|}}$$