Assignment (EECS6327 F16)

Due: in class on Oct 21, 2016.

You have to work individually. Hand in a hardcopy of your answers before the deadline. No late submission will be accepted. No handwritting is accepted. Direct your queries to Hui Jiang (hj@cse.yorku.ca).

1. **Mutual Information**: Assume we have a random vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ which follows a bivariate Gaussian distribution: $\mathcal{N}(\mathbf{x}|\mu, \Sigma)$, where $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ is the mean vector and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma 2 \\ \rho\sigma_1\sigma 2 & \sigma_2^2 \end{pmatrix}$ is the covariance matrix. Derive the formula to compute mutual information between x_1 and x_2, i.e., $I(x_1, x_2)$.

   Hints: May need to use the formula: $\Gamma(n+1) = n \cdot \Gamma(n)$.

2. **KL Divergence**: Assume we have two multi-variate Gaussian distributions: $\mathcal{N}(\mathbf{x}|\mu_\mathbf{1}, \mathbf{\Sigma_1})$ and $\mathcal{N}(\mathbf{x}|\mu_\mathbf{2}, \mathbf{\Sigma_2})$, where $\mu_\mathbf{1}$ and $\mu_\mathbf{2}$ are their mean vectors, and $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ are their covariance matrices. Derive the formula to computer the KL divergence between these two Gaussian distributions.

3. **Classification with rejection**: In many pattern classification problems, one has the option either to assign the pattern to one of N classes, or to *reject* it as being unrecognizable. If the cost for rejection is not too high, rejection may be a desirable action. If we observe feature $\mathbf{x}$ of a pattern (assume its true class id is ω_i), let's define the loss function for all actions α_j as:

$$\lambda(\alpha_j|\omega_i) = \begin{cases} 0 & : \quad j = i \ \text{(correct classification)} \\ \lambda_s & : \quad j \neq i \ \text{and} \ 1 \leq j \leq N \text{(wrong classification)} \\ \lambda_r & : \quad \text{rejection} \end{cases}$$

   where λ_s is the loss incurred for making any a wrong classification decision, and λ_r is the loss incurred for choosing the rejection action. Show the minimum risk is obtained by the following decision rule: we decide ω_i if $p(\omega_i|\mathbf{x}) \geq p(\omega_j|\mathbf{x})$ for all j and if $p(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

4. **Linear-Gaussian models**: Consider a joint distribution $p(\mathbf{x}, \mathbf{y})$ defined by the marginal and conditional distributions as follows:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \mu, \mathbf{\Delta^{-1}})$$

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y} \mid \mathbf{Ax} + \mathbf{b}, \mathbf{L^{-1}}),$$

derive and find expressions for the mean and covariance of the marginal distribution $p(\mathbf{y})$ in which the variable $\mathbf{x}$ has been integrated out.

Hints: You may need to use the Woodbury matrix inversion formula:

$(\mathbf{A} + \mathbf{ABC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$.

5. **Discriminant Analysis**: Let $(\mathbf{x}, y) \in \mathcal{R}^d \times \{0, 1\}$ be a random pair such that $\Pr(y = k) = \pi_k > 0$ $(\pi_0 + \pi_1 = 1)$ and the conditional distribution of $\mathbf{x}$ given y is $p(\mathbf{x} \mid y) = \mathcal{N}(\mathbf{x} \mid \mu_y, \Sigma_y)$, where $\mu_0 \neq \mu_1 \in \mathcal{R}^d$ and $\Sigma_0, \Sigma_1 \in \mathcal{R}^{d \times d}$ are mean vectors and covariance matrices respectively.

   (a) What is the (unconditional) density of $\mathbf{x}$?

   (b) Assume that $\Sigma_0 = \Sigma_1 = \Sigma$ is a positive definite matrix. Compute the Bayes classifier. What is the nature of separation boundary between two classes?

   (c) Assume that $\Sigma_0 \neq \Sigma_1$ are two positive definite matrices. Compute the Bayes classifier. What is the nature of separation boundary between two classes?