Assignment One (EECS6327 F19)

Due: in class on Oct 2, 2019.

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

1. **Multinomial** vs. **Dirichlet**:

   (a) Given a Multinomial distribution of m discrete random variables:

   $$\begin{aligned} & \Pr(X_1 = r_1, X_2 = r_2, \cdots, X_m = r_m \mid p_1, p_2, \cdots, p_m) \\ = & \frac{(r_1 + \cdots + r_m)!}{r_1! \cdots r_m!} p_1^{r_1} \times p_2^{r_2} \times \cdots \times p_m^{r_m} \end{aligned} \tag{1}$$

   where $X_1, \cdots, X_m$ take all non-negative intergers $r_1 \geq 0, \cdots, r_m \geq 0$ that satisfies $\sum_{i=1}^{m} r_i = N$. Prove that the multinomial distribution satisfies the sum-to-one normalization constraint:

   $$\sum_{X_1, \cdots, X_m} \Pr(X_1 = r_1, X_2 = r_2, \cdots, X_m = r_m \mid p_1, p_2, \cdots, p_m) = 1.$$

   (b) Given a Dirichlet distribution of m continuous random variables:

   $$\begin{aligned} & \Pr(X_1 = p_1, X_2 = p_2, \cdots, X_m = p_m \mid r_1, r_2, \cdots, r_m) \\ = & \frac{\Gamma(r_1 + \cdots + r_m)}{\Gamma(r_1) \cdots \Gamma(r_m)} p_1^{r_1 - 1} \times p_2^{r_2 - 1} \times \cdots \times p_m^{r_m - 1}, \end{aligned} \tag{2}$$

   derive the following results for the mean and variance:

   $$\mathbb{E}(X_i) = \frac{r_i}{r_0}$$

   $$\mathrm{Var}(X_i) = \frac{r_i(r_0 - r_i)}{r_0^2(r_0 + 1)}$$

   where we denote $r_0 = \sum_{i=1}^{m} r_i$.

   Hints: $\Gamma(x+1) = x \cdot \Gamma(x)$.

2. **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., $I(x_1, x_2)$.

3. **KL Divergence**: Assume we have two multi-variate Gaussian distributions: $\mathcal{N}(\mathbf{x}|\mu_1, \mathbf{\Sigma}_1)$ and $\mathcal{N}(\mathbf{x}|\mu_2, \mathbf{\Sigma}_2)$, where μ_1 and μ_2 are their mean vectors, and $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are their covariance matrices. Derive the formula to compute the KL divergence between these two Gaussian distributions.

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

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

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

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:

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

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?