CS 215 : Data Analysis and Interpretation

(Instructor : Suyash P. Awate)

Quiz (Closed Book)

Roll Number: ____

Name: ____

For all questions, if you feel that some information is missing, make justifiable assumptions, state them clearly, and answer the question.

Relevant Formulae

• Poisson: $P(k|\lambda) := \lambda^k \exp(-\lambda)/(k!)$

• Exponential: $P(x; \lambda) = \lambda \exp(-\lambda x); \forall x > 0$

• Gamma:

$$P(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}$$

• Gamma function: $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ for real-valued z.
When z is integer valued, then $\Gamma(z) = (z-1)!$, where ! denotes factorial.
For all z, $\Gamma(z+1) = z\Gamma(z)$.

• Univariate Gaussian:

$$P(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-0.5 \frac{(x-\mu)^2}{\sigma^2}\right)$$

• Product of two univariate Gaussians: $G(z; \mu_1, \sigma_1^2) G(z; \mu_2, \sigma_2^2) \propto G(z; \mu_3, \sigma_3^2)$
where

$$\mu_3 = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}; \sigma_3^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

• Multivariate Gaussian:

$$G(x; \mu, C) = \frac{1}{(2\pi)^{D/2} |C|^{0.5}} \exp(-0.5 (x-\mu)^\top C^{-1} (x-\mu))$$

• $d(Ax) = A dx$
• $d(x^\top A x) = x^\top (A + A^\top) dx$

---

1. (10 points)

   Consider a univariate continuous random variable X that takes values on the positive real line.

   If $P(X)$ is a uniform (improper) prior PDF on the positive real line, then derive the PDF for the random variable $Y := X^2$.

   If $P(A)$ is a uniform (improper) prior PDF on the positive real line, then derive the PDF for the random variable $B := \sqrt{A}$.

   Follow the process covered in transformation of random variables.

   $P(X) \propto 1$ gives $P(Y) \propto 1/\sqrt{Y}$

   $P(A) \propto 1$ gives $P(B) \propto B$

   ---

2. (10 points)

   Let X be a univariate Gaussian with zero mean and standard deviation σ.

   Let Y be a univariate Gaussian with zero mean and standard deviation σ.

   If X and Y are independent, then what is the PDF of random variable $Z := Y/X$? Derive theoretically.

   ---

   http://www.math.wm.edu/~leemis/chart/UDR/PDFs/StandardnormalStandardcauchy.pdf

   In our question, the Gaussians have variance σ^2, unlike the case at the link where the variance is one. However, this doesn't change the answer, because the random variables X and Y can be modeled as $\sigma X'$ and $\sigma Y'$ where X' and Y' are standard normal.

   ---

3. (10 points)

   Consider independent random vectors X and Y that both have D-variate Gaussian probability density functions (PDFs), each with some arbitrary mean and some arbitrary covariance matrix.

   Does the random vector $Z := X - Y$ also have a D-variate Gaussian PDF ? If so, prove why so. If not, prove why not.

   In either case, find (mathematically derive) the mean and the covariance matrix of Z, in terms of the parameters underlying $P(X)$ and $P(Y)$; do it for the most general case.

   Let $X := A_1 W_1 + b_1$ (by definition), with mean b_1 and covariance matrix $C_1 = A_1 A_1^\top$, where A_1 is a square matrix

   Let $-Y := A_2 W_2 + b_2$ (by definition), with mean b_2 and covariance matrix $C_2 = A_2 A_2^\top$, where A_2 is a square matrix

   Then $X - Y = [A_1; A_2][W_1^\top; W_2^\top]^\top + (b_1 + b_2)$, where $[A_1; A_2]$ is a rectangular matrix of size $D \times 2D$

   Thus, $X - Y$ is also multivariate Gaussian (by definition)

   Thus, $X - Y$ has mean $b_1 + b_2$ and covariance matrix $[A_1; A_2][A_1; A_2]^\top = A_1 A_1^\top + A_2 A_2^\top = C_1 + C_2$

   ---

4. (10 points)

   Consider that you are performing maximum-likelihood classification for data in a D-dimensional space. There are two classes each having a multivariate Gaussian PDF.

   The first class has the PDF given by $G_1(X; \mu, C_1)$ where $C_1 := \sigma_1^2 I$, where I is the identity matrix and σ_1 is a scalar.

   The second class has the PDF given by $G_2(X; \mu, C_2)$ where $C_2 := \sigma_2^2 I$, where I is the identity matrix and σ_2 is a scalar.

   If $0 < \sigma_1 < \sigma_2 < \infty$, then find the region in the D-dimensional space where data points will get classified into the first class.

   A hypersphere with center as the origin and the squared-radius as:

   $D \log(\sigma_2/\sigma_1)/(0.5/\sigma_1^2 - 0.5/\sigma_2^2)$

   ---

5. (10 points)

   Suppose you have a dataset with N observations $\{x_i \in \mathbb{R}^{10}\}_{i=1}^{N}$. Suppose you want to visualize the dataset as a scatter plot in three dimensions. Of course, that 3-D visualization cannot always be an exact representation of the original 10-D dataset, but you would like to create a representation of the data, for the purpose of visualization, that is as good as possible (based on the concepts we covered in class). Provide an (implementable) algorithm that takes the original dataset and produces a representative dataset $\{y_i \in \mathbb{R}^3\}_{i=1}^{N}$ in 3-D that is easily visualizable.

   Given an algorithm for dimensionality reduction using PCA. As discussed in class. As in the assignment.

   ---

6. (15 points)

   Consider a system where a multivariate input vector X (of size $N \times 1$) gets transformed through a matrix A (of size $M \times N$, where $M < N$) to lead to an output vector B (of size $M \times 1$). Thus, $B = AX$.

   The input vector X is unknown (unobserved).

   We try to measure B, but because our measurement system is erroneous, our measurement reflects a corrupted version of B. Let this measurement, i.e., the observed data, be vector C, where each element in the vector C undergoes an independent additive standard-normal perturbation over the corresponding element in the vector B.

   • [5 points] Given A and C, formulate the estimation of X as a maximum-likelihood (ML) estimation (optimization) problem. Will this ML-estimation problem have a unique solution ? If so, first prove that it will have a unique solution and then give (mathematically derive) an algorithm to find that solution. If not, prove/argue why not ?

   Data = C

   Likelihood: $P(C|X) = G(C; AX, \sigma^2 I)$, where $\sigma = 1$

   ML-estimation formulation: $\arg\min_x 0.5\|C - Ax\|_2^2/\sigma^2$

Because A is singular (because $M < N$), there will be multiple values of x for which $C = Ax$

• [5 points] Assume a prior model on X that indicates that X is drawn from a multivariate Gaussian with zero mean and a diagonal covariance matrix λI, where real-valued $\lambda > 0$ and I is the $N \times N$ identity matrix.

Given A and C and the prior model, formulate the estimation of X as a maximum-a-posterior (MAP) estimation (optimization) problem.

Data = C

Likelihood: $P(C|X) = G(C; AX, \sigma^2 I)$, where $\sigma = 1$

Prior: $P(X) = G(X; 0, \lambda I)$

MAP-estimation formulation: $\arg\min_x 0.5\|C - Ax\|_2^2/\sigma^2 + 0.5\|x\|_2^2/\lambda$

Taking the derivative of the objective function with respect to x and assigning it to zero gives:

$A^\top C/\sigma^2 = (A^\top A/\sigma^2 + I/\lambda)x$

• [5 points] Will this estimation problem have a unique solution ? If so, first prove that it will have a unique solution and then give (mathematically derive) an algorithm to find that solution. If not, prove/argue why not ?

---

(1)

We can prove that $A^\top A$ is a symmetric positive semi-definite matrix

(2)

Thus, $A^\top A$ has an eigen decomposition of the form $QDQ^\top$, where elements on diagonal of D are non-negative

Thus, $A^\top A/\sigma^2 + I/\lambda$ has en eigen decomposition of the form $Q(D/\sigma^2 + I/\lambda)Q^\top$, where elements on the diagonal $(D/\sigma^2 + I/\lambda)$ are all positive

Thus, we proved that $A^\top A/\sigma^2 + I/\lambda$ is invertible.

(3)

Thus, the optimization problem will have a unique solution given by:

$x^* = (Q(D/\sigma^2 + I/\lambda)^{-1}Q^\top)(A^\top c/\sigma^2)$
