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:

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

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

---

1. [15 points]

   Consider the 2×2 matrix A where the element $A_{2,2} = 0$ and all other elements are equal to 1.

   • (5 points) Formulate the problem of finding all singular values and singular vectors of this matrix A as one or more optimization problems.

   ---

   see class notes for how to define the principal right-singular vector v_1 and value σ_1, using the matrix norm

   principal left-singular vector $u_1 := Av_1/\|Av_1\|_2$, and $\sigma_1 := \|Av_1\|_2$

   also see how to define the other singular vectors and values. In our 2D case, the second singular vector is simply a unit-norm vector orthogonal to the first.

   second left-singular vector $u_2 := Av_2/\|Av_2\|_2$, and $\sigma_2 := \|Av_2\|_2$

   • (5 points) For the largest singular value, write pseudocode to plot the objective function as a function of its (discretized) argument. Write pseudocode to find its (approximate) optimal solution using a search over the discretized argument.

   ---

   see class notes.

   let a unit vector in 2D be $[\cos(\theta), \sin(\theta)]^\top = [c, s]^\top$

   then, $A[c, s]^\top = [c + s, c]^\top$

   objective function: $(c^2 + s^2 + 2cs) + c^2 = 1 + c^2 + 2cs$

   A = [[1 1] ; [1 0]], [U S V] = svd (A)
   t = [0:0.01:pi]; c=cos(t); s=sin(t); f = sqrt((c+s).^2 + c.^2);
   plot (t,f,'ro-'), grid on, axis tight, pause, close
   [val ind] = max(f), c(ind),s(ind)
   v1 = [c(ind), s(ind)]$^\top$, sigma1 = norm(A*v1), u1 = A*v1 / norm(A*v1)
   v2 = [s(ind), -c(ind)]$^\top$, sigma2 = norm(A*v2), u2 = A*v2 / norm(A*v2)

   • (5 points) Argue whether the largest singular value will be $> \sqrt{2}$, or $\geq \sqrt{2}$, or within $[1, \sqrt{2}]$, or within $(1, \sqrt{2})$, or ≤ 1 or < 1 ? Give clear mathematical justifications.

   ---

   see class notes.

   must be $\geq \sqrt{2}$, because:
   for unit-norm vector $x = [1, 0]^\top$, we get $Ax = [1, 1]^\top$ that has norm sqrt2 (2 points)

   must be $> \sqrt{2}$, because:
   slope of objective function at $\theta = 0$ is
   $-2\cos(\theta)\sin(\theta) + 2(\cos(\theta))^2 - 2(\sin(\theta))^2$
   > 0 (3 points)

   ---

2. [12 points]

   Consider a real-valued symmetric positive-definite matrix A of size D×D, with the largest eigenvalue being equal to 1 and being unique (other eigenvalues needn't be unique).

   Consider a random unit-norm vector $x \in \mathbb{R}^D$.

   Consider the sequence of vectors $y_n := A^n x$ for $n = 1, 2, 3, \cdots$, such that $y_1 := Ax$, $y_2 := AAx$, etc.

   What can you infer about the vector $\lim_{n\to\infty} y_n$, whatever x may be ? Give a clear mathematical justification.

Consider the eigen decomposition $A = V\Lambda V^\top$, where eigenvalues are sorted in decreasing order on the diagonal

Let $x = Vz$

$A^n x = (V\Lambda^n V^\top)(Vz) = V\Lambda^n z$

If largest eigenvalue is 1, then $\lim_{n\to\infty} \Lambda^n$ is a diagonal matrix with only first diagonal element as 1, and rest all zero. (2 points)

Case 1: x isn't orthogonal to principal eigenvector v_1 (5 points). Then, $\Lambda^n z$ is a non-zero vector with only first component as non-zero. So, $\lim_{n\to\infty} y^n$ is parallel to the principal eigenvector.

Case 2: x is orthogonal to principal eigenvector v_1 (5 points). Then, $\Lambda^n z$ is a zero vector. So, $\lim_{n\to\infty} y^n$ is a zero vector.

---

3. [10 points] Consider a dataset $\{x_n \in \mathbb{R}^D\}_{n=1}^N$.

   • (5 points) Mathematically derive the relationship between the covariance of the given dataset and the covariance of another dataset $\{y_n := Ax_n \in \mathbb{R}^D\}_{n=1}^N$, where A is a D$\times$D invertible matrix.

   ---

   please see lecture notes.

   • (5 points) Consider a uniform distribution on a unit square within the 2D Euclidean plane; the square is centered at the origin and has its sides parallel to the cardinal axes. Consider re-representing this distribution in an arbitrarily shifted and arbitrarily rotated coordinate frame, e.g., through an appropriate transformation of the random variables.

   What will the covariance matrix of the re-represented distribution be (in the new coordinate frame) ? Give a clear mathematical justification.

   What will be the eigenvalues and eigenvectors of the new covariance matrix ?

   ---

   please see lecture notes.

   ---

4. [18 points] Consider a Poisson distribution with parameter $\lambda > 0$. Consider a prior on λ as the Gamma distribution with parameters α, β.

• (5 points) For a dataset with sample size N, where each datum is independently drawn from a Poisson distribution, derive a closed-form mathematical expression for the posterior distribution on λ.

---

see lecture notes

posterior PDF is also a Gamma PDF with parameters $(\sum_n k_n + \alpha, n + \beta)$

• (5 points) Derive (from scratch, without using any known result) a closed-form mathematical expression for the mode of the posterior distribution.

---

derivation for the mode of Gamma PDF is here:
`https://en.wikipedia.org/wiki/Gamma_distribution`
derive this by maximizing the PDF function, using appropriate arguments for the two cases

use these expressions to derive the mode of the posterior

• (5 points) Derive (from scratch, without using any known result) a closed-form mathematical expression for the mean of the posterior distribution.

---

derivation for the mean of Gamma PDF is here:
`https://proofwiki.org/wiki/Expectation_of_Gamma_Distribution`
derive it

use these expressions to derive the mode of the posterior

• (3 points) As sample size of the data tends to infinity, what does the posterior mean tend to ? Give a mathematical derivation.

---

see lecture notes. it tends to the MLE, i.e., the sample mean of the observed data sample

---