Problem 1: Statistics Review (20 points)

1. **(10 points)** Recall the Dirichlet distribution, a family of continuous multivariate probability distribution used as prior to the multinomial distribution, with the PDF specified as:

$$f(\mathbf{x};\alpha) = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K} x_i^{\alpha_i - 1},$$

where $\Gamma(x)$ is the Gamma function with derivative $\Gamma(x)' = \phi(x)$, the digamma function. Derive the maximum likelihood estimator for parameter α for the Dirichlet distribution.

2. **(10 points)** Use the *pdf kernel method* to derive the posterior distribution the following conjugate pair: prior $p(\theta) = \mathcal{N}(\theta; 0, I)$ and likelihood function $p(x;\theta) = \mathcal{N}(x;\mu,\Sigma)$. Show all steps.

Answer: Q1

To derive the maximum likelihood estimator (MLE) for the parameter vector "α" in the Dirichlet distribution, need to maximize the log-likelihood function. The log-likelihood function for the Dirichlet distribution is given by:

$$\mathrm{L}(\alpha) = \log(\mathrm{x};\alpha) = \log\Gamma\sum_{i=1}^{K} a_i - \sum_{i=1}^{K} log\Gamma(a_i) + \sum_{i=1}^{K}(a_i - 1)\log(x_i)$$

To find the MLE, need to differentiate the log-likelihood function with respect to each parameter a_i and set the derivatives equal to zero:

$$\frac{\partial L(\alpha)}{\partial \alpha_i} = \frac{\partial(\log\left(\Gamma\sum_{i=1}^{K}\alpha_i\right) - \sum_{i=1}^{K} log\Gamma(\alpha_i) + \sum_{i=1}^{K}(\alpha_i - 1)\log(x_i))}{\partial \alpha_i}$$

$$= \phi\left(\sum_{i=1}^{K}\alpha_i\right) - \phi(\alpha_i) + \log(\alpha_i) = 0$$

where $\phi(\mathrm{x})$ is the digamma function, the derivative of the log of the gamma function. Rearranging the equation, obtain:

$$\phi(\alpha_\mathrm{i}) - \phi\left(\sum_{i=1}^{K}\alpha_i\right) = \log(\alpha_i)$$

This equation does not have a closed-form solution, so typically solve it iteratively using numerical methods such as Newton's method or gradient descent. The iterative procedure involves starting with an initial guess for α and updating it until convergence is achieved.

During each iteration, compute the digamma function values and the log of the data points. Then, update the parameter vector α using the equation:

$$\alpha_i^{t+1} = \alpha_i^t + \phi_i^t - \phi\left(\sum_{i=1}^{K}\alpha_i^t\right) + \log(x_i)$$

where t is the iteration number.

repeatting this iterative process until the parameter estimates converge. The resulting parameter vector α will be the maximum likelihood estimator for the Dirichlet distribution.

Answer: Q2

To derive the posterior distribution using the PDF kernel method for the conjugate pair consisting of a Gaussian prior and Gaussian likelihood, follow these steps:

1. Prior Distribution: $\mathrm{p}(\theta) = \mathrm{N}(\theta; 0, \mathrm{I})$

   This is a multivariate Gaussian distribution with mean 0 and identity covariance matrix I.

2. Likelihood Function: $p(x|\theta) = N(x; \mu, \sigma)$

   This is a multivariate Gaussian distribution with mean μ and covariance matrix σ.

3. Compute the joint distribution: : $p(x, \theta) = p(x|\theta) * p(\theta)$

   Since the prior and likelihood are independent, can multiply them to obtain the joint distribution.

$$p(x, \theta) = (x; \mu, \sigma) * N(\theta; 0, I)$$

4. Apply the pdf kernel method:

   use the properties of Gaussian distributions to simplify the joint distribution.

$$p(x, \theta) = \frac{1}{(2\pi)^{d/2} \times |\sigma|^{0.5}} \exp(-0.5 \times (x-\mu)^T \times \sigma^{-1} \times (x-\mu)) \times \frac{1}{(2\pi)^{\frac{k}{2}} \times |I|^{0.5}} \exp(-0.5 \times \theta^T * \theta)$$

   where *d* is the dimensionality of x, and *k* is the dimensionality of θ.

5. Simplify the joint distribution: Combining the terms, obtain:

$$\begin{aligned} p(x, \theta) &= \frac{1}{(2\pi)^{d/2} \times |\sigma|^{0.5} \times (2\pi)^{\frac{k}{2}} \times |I|^{0.5}} \exp(-0.5 \times (x-\mu)^T \times \sigma^{-1} \times (x-\mu) - 0.5 \times \theta^T * \theta) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}+\frac{k}{2}} \times |\sigma|^{0.5} \times |I|^{0.5}} \exp(-0.5 \times (x-\mu)^T \times \sigma^{-1} \times (x-\mu) - 0.5 \times \theta^T * \theta) \end{aligned}$$

6. Compute the posterior distribution:

   To obtain the posterior distribution, need to normalize the joint distribution by dividing it by the marginal likelihood, which acts as a normalization constant. The marginal likelihood can be obtained by integrating the joint distribution over all possible values of θ.

$$p(\theta|x) = \frac{p(x, \theta)}{\int p(x, \theta) d\theta}$$

   The denominator is the normalization constant that ensures the posterior distribution integrates to 1.

   By simplifying and normalizing the joint distribution, obtain the posterior distribution.