Name: Luke Trujillo
Due Date: February 5, 2020

143 Hw #1

Problem 1 Read and write up Matrix Normal Distribution in your own words.

Solution: When one encounters the concept of the probability distribution, they encounter the canonical example of the normal distribution; a special probability distribution for a single random variable x. Given a mean μ and standard deviation σ, the relationship is denoted as $x \sim \mathcal{N}(\mu, \sigma^2)$, and the density function is given by

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

We can go even further and consider the case where we have a set of random variables $x_1, x_2, \ldots, x_n$ which we can assemble into a vector $\boldsymbol{x}$. We then say our vector $\boldsymbol{x}$ has a multivariate distribution if and only if for each component x_i we have $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for some means μ_i and standard deviations σ_i. In this case, we write $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and the density function is given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

where $\boldsymbol{\mu}$ is the vector of mean values μ_i and Σ is the $n \times n$ covariance matrix with values $\Sigma_{i,j} = E[(x_i - \mu_i)(x_j - \mu_j)]$.

Finally, we can generalize this even further if we have a $n \times p$ matrix X of random values. To make a logical definition, we want to be able to fall back on our previous definition of the multivariate normal distribution. Therefore let $\text{vec}(X)$ be the flattening of a matrix, M be the $n \times p$ mean value matrix, and $\Sigma \otimes \Psi$ be the covariance matrix, where Σ is $n \times n$ and Ψ is $p \times p$, and where $\otimes$ is the Kronecker product. Then X is distributed via the matrix normal distribution if and only if $\text{vec}(X) \sim \mathcal{MN}(\text{vec}(M), \Sigma \otimes \Psi)$ is distributed via the multivariate normal distribution. We then write the density function as

$$f(X) = \frac{1}{(2\pi)^{np/2} \det(\Sigma)^{p/2} \det(\Psi)^{n/2}} \exp\left(-\frac{1}{2}\text{tr}(\Sigma^{-1}(X-M)^T \Psi^{-1}(X-M))\right).$$

One can show that our definition is consistent as follows. Suppose that $\text{vec}(X)$ has a multivariate distribution. Then its probability density function must be

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^{np} \det(\Sigma \otimes \Psi)}} \exp\left(-\frac{1}{2}(\text{vec}(X) - \text{vec}(M))^T (\Sigma \otimes \Psi)^{-1}(\text{vec}(X) - \text{vec}(M))\right)$$

which we can reduce to

$$\frac{1}{\sqrt{(2\pi)^{np}\det(\Sigma)^p\det(\Psi)^n}}\exp\left(-\frac{1}{2}(\mathrm{vec}(X-M))^T(\Sigma^{-1}\otimes\Psi^{-1})\mathrm{vec}(X-M)\right)$$
$$=\frac{1}{(2\pi)^{np/2}\det(\Sigma)^{p/2}\det(\Psi)^{n/2}}\exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(X-M)^T\Psi^{-1}(X-M)\right)$$

via Kronecker properties on determinants and inverses. Hence the definition and probability density function makes sense.

□

Problem 2 Read and write up Dirichlet distribution and find an example using Dirichlet (e.g. LDA).

Solution: The Dirichlet distribution is a type of multivariate probability distribution for a random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_K)$ governed by parameters $\alpha_1, \alpha_2, \ldots \alpha_K > 0$ which we may assemble into a vector $\boldsymbol{\alpha}$. For this distribution, we must assume that our random variables $x_i \geq 0$ and sum to one. Since we have K-many elements which make up $\boldsymbol{x}$, one can imagine that it must live on a simplex; the fact that they sum to one forces them to live on the $K-1$ simplex.

In this case, the density function is given by

$$f(\boldsymbol{x}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

where $B(\boldsymbol{\alpha})$ is the beta function

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K}\Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)}$$

with Γ being the gamma function.

It is true that the Dirichlet distribution is applied extensively in natural language processing, specifically in categorizing text in a written document by topic. Previously, the multinomial distribution had been used to handle the problem of topic analysis, but researchers identified a "bursting" phenomenon which it failed to accurately model. This is simply the idea that, when a word shows up, it has a high chance of showing up again shortly after, and this is key to understanding how documents are written. The Dirichlet distribution was proposed as a placement, which takes each document and represents it as a probability vector (where the probability is of a word appearing).

A different application can be found in imagine analysis, which is similar to the problem handled by LDA and discussed in the paper *Unsupervised Learning of a Finite Mixture Model Based on the Dirichlet Distribution and Its Application*. In the paper, the authors discuss the benefits of combining both the multinomial and Dirichlet distributions, as well as applications in image searching (e.g. using an image to find other related images). They also classified images with respect to skin color, and skin identification in general.

□

Problem 3 Find an example of manifolds which Prof. Gu has not yet mentioned before.

Solution: In class, we've discussed real projective spaces and showed how they are a manifold since they are the quotient of manifolds. However, there are also complex projective spaces, and in looking for more manifolds, I learned about the quarternionic projective space $\mathbb{HP}^n$, which is also an example of a manifold. As one might expect, it acts on the quarternions $a + bi + cj + dk$. It is the generalization of a complex projective space, just as the complex projective space is the generalization of the real projective space, although they don't appear to be well studied.

Its construction is no different from the construction of the real projective space or the complex projective space. We defined an equivalence relation on quarternios $(q_0, q_1, \ldots, q_{n-1})$ where two tuples are equivalent if one is a scalar multiple of the other. We then consider the group of units $G = H^\times$, and define $\mathbb{HP}^n$ to be the orbit of the group action of G on $\mathbb{H}^{n+1}/\{(0, 0, \ldots, 0)\}$. Since this is therefore a homogeneous space, we see that it is a manifold.

□