Andres Fernandez – 5692442 – fr_andres@msn.com

Exercise 1

Consider a multivariate normal distribution in variable x with mean μ and covariance Σ. Show that if we make the linear transformation $y = Ax + b$ then the transformed variable y is distributed as:

$$p(y) = \mathcal{N}(A\mu + b, A\Sigma A^T)$$

i)

Asked is then to show that the normal distribution of an affine transformation is a normal distribution itself, with the mentioned transformations applied to its parameters. For that sake it is necessary to remember the definition of the multivariate normal function (for a positive, semi-definite covariance matrix Σ):

$$\mathcal{N}(\vec{x}|\vec{\mu}, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})\right) \tag{1}$$

And substituting it with the corresponding affine transformation yields the following formula (the vector arrows will be from now on omitted):

$$\mathcal{N}(Ax + b|\mu, \Sigma) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}((Ax + b) - \mu)^T \Sigma^{-1} ((Ax + b) - \mu)\right) \tag{2}$$

ii)

In order to reduce equation 2, the following properties of linear algebra are assumed:

$$\begin{aligned}
&(A^{-1})^T = (A^T)^{-1} = AA^{-T} \\
&(AB)^T = B^T A^T \\
&(AB)^{-1} = B^{-1} A^{-1} \\
&A^{-1}A = A^T A^{-T} = I \\
&IA = AI = A \\
&y = Ax + b \iff x = A^{-1}(y - b)
\end{aligned} \tag{3}$$

This assumptions are valid for a matrix A of any size, having the identity matrix I the corresponding number of dimensions in each case.

iii)

Now it's possible to perform the transformation in the body of the exponential function:

$$\begin{gathered}(x-\mu)^T\Sigma^{-1}(x-\mu) = \\ ((A^{-1}(y-b))-\mu)^T\Sigma^{-1}((A^{-1}(y-b))-\mu) = \\ ((A^{-1}(y-b))-\mu)^T(\mathbf{A^T}A^{-T})\Sigma^{-1}(\mathbf{A^{-1}A})((A^{-1}(y-b))-\mu) = \\ \left(((A^{-1}(y-b))-\mu)^T A^T\right)\left(\mathbf{A^{-T}\Sigma^{-1}A^{-1}}\right)\left(A((A^{-1}(y-b))-\mu)\right) = \\ \left(\mathbf{A}(\mathbf{A^{-1}}(\mathbf{y}-\mathbf{b}))-\mu\right)^{\mathbf{T}}\left(A^{-T}\Sigma^{-1}A^{-1}\right)\left((\mathbf{AA^{-1}}(\mathbf{y}-\mathbf{b}))-\mathbf{A}\mu\right) = \\ (\mathbf{y}-\mathbf{b}-\mathbf{A}\mu)^{\mathbf{T}}\left(A^{-T}\Sigma^{-1}A^{-1}\right)(\mathbf{y}-\mathbf{b}-\mathbf{A}\mu) = \\ (y-(\mathbf{A}\mu+\mathbf{b}))^T(\mathbf{A\Sigma A^T})^{-\mathbf{1}}(y-(\mathbf{A}\mu+\mathbf{b}))\end{gathered} \tag{4}$$

Whereas $A^{-T}\Sigma^{-1}A^{-1} = A^{-T}(A\Sigma)^{-1} = A\Sigma A^{T-1}$. Also, recall that the normalization factor remains the same regardless of the dimensions of A, since it depends only on the original Σ. Therefore, the following equivalence holds:

$$\begin{aligned}\rho(x) = \rho(A^{-1}(y-b)) =& \frac{1}{\sqrt{|2\pi\Sigma|}}\exp\left(-\frac{1}{2}(y-(\mathbf{A}\mu+\mathbf{b}))^T(\mathbf{A\Sigma A^T})^{-\mathbf{1}}(y-(\mathbf{A}\mu+\mathbf{b}))\right) = \\ &= \mathcal{N}(y|A\mu+b, A\Sigma A^T)\end{aligned} \tag{5}$$

□

Exercise 2

Show that we can convert a normal distribution with mean μ and covariance Σ to a new distribution with mean 0 and covariance I using the linear transformation $y = Ax + b$ where

$$A = \Sigma^{-\frac{1}{2}}$$
$$b = -\Sigma^{-\frac{1}{2}}\mu$$

This is known as the whitening transform. Note that M is the square root of a matrix Q, i.e. $M = Q^{\frac{1}{2}}$, if $MM = Q$

i)

Here, the following linear algebra assumptions are needed:

$$\begin{aligned}
&I \cdot I = I \iff I^{-1} = I \\
&A^{-\frac{1}{2}} A^{-\frac{1}{2}} = A^{-1} \\
&A^{-\frac{1}{2}} A^{\frac{1}{2}} = I \\
&\Sigma = \Sigma^T \iff \Sigma^{-1} = \Sigma^{-T} \iff \Sigma^{\frac{1}{2}} = (\Sigma^{\frac{1}{2}})^T \\
&A^{-\frac{1}{2}} A^{\frac{1}{2}} = I
\end{aligned} \tag{6}$$

And also the following implication: $y = Ax + b \iff x = A^{-1}(y - b)$, which, for the given values, yields:

$$x = (\Sigma^{-\frac{1}{2}})^{-1}(y - (-\Sigma^{-\frac{1}{2}}\mu)) = \Sigma^{\frac{1}{2}}(y + \Sigma^{-\frac{1}{2}}\mu) = \Sigma^{\frac{1}{2}} y + \mu$$

ii)

Now, substituting in the exponential body of the normal PDF,

$$\begin{aligned}
&\exp\left(-\frac{1}{2}((\Sigma^{\frac{1}{2}} y + \mu) - \mu)^T \Sigma^{-1}((\Sigma^{\frac{1}{2}} y + \mu) - \mu)\right) = \\
&\exp\left(-\frac{1}{2}(\Sigma^{\frac{1}{2}} y)^T \Sigma^{-1}(\Sigma^{\frac{1}{2}} y)\right) = \\
&\exp\left(-\frac{1}{2}(\Sigma^{\frac{1}{2}} y)^T \Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}}(\Sigma^{\frac{1}{2}} y)\right) = \\
&\exp\left(-\frac{1}{2}(y^T (\Sigma^{\frac{1}{2}})^T \Sigma^{-\frac{1}{2}}(\Sigma^{-\frac{1}{2}} \Sigma^{\frac{1}{2}} y)\right) = \\
&\exp\left(-\frac{1}{2}(y^T y)\right) = \exp\left(-\frac{1}{2}((y - 0)^T I (y - 0))\right)
\end{aligned} \tag{7}$$

Which yields a normal distribution with zero value for μ and the identity matrix for Σ.

Exercise 3

Now, we want to exploit the knowledge acquired in Exercises 1 and 2 to conceive a method for sampling random vectors from an arbitrary multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. Sampling from a distribution means, the probability of drawing a given vector is proportional to the pdf of this distribution. First, we sample a vector of length N (number of dimensions) from a normal distribution $\mathcal{N}(0, I)$. Since all components are uncorrelated, we can easily do this by drawing N random numbers using a built-in random number generator for univariate normal distributions. Then we transform the samples *with appropriate matrix B, such that the new distribution has the desired covariance. Finally, we shift the samples to the desired mean value v.*

a) How are B, Σ and v related?

If I understood correctly, the described setup intends to achieve the transformation from a standard normal multivariate distribution $x_1 \sim \mathcal{N}(0, I)$ to any other multivariate distribution $x_2 \sim \mathcal{N}(\mu, \Sigma)$, by application of the affine transformation $y = Bx + v$. Based on the relations showed before, we know that $y \sim \mathcal{N}(B \cdot 0 + v, B \cdot I \cdot B^T) = \mathcal{N}(v, BB^T)$. Therefore,

v becomes directly the new expected value, that is: $v = \mu$

BB^T becomes the new covariance matrix Σ.

If B is a triangular and square matrix, this is the **Cholesky decomposition**.

b) Sample 1000 values from the two-dimensional normal distribution with mean vector $\mu = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, and the covariance matrix $\Sigma = \begin{pmatrix} 4 & -0.5 \\ -0.5 & 2 \end{pmatrix}$. Plot a two-dimensional histogram and explain why this plot shows approximately the desired distribution. Use $M \times M$ quadratic bins of equal size. The value of a given bin is defined as the number of samples that fall into this bin, divided by the total number of samples. To plot the histogram, represent bin values by grey values (or color) and plot these grey values into a two-dimensional coordinate system (like an 'image')

See Figure 2, and see/run the Python2 script *fernandez_blatt3.py* for the details. This

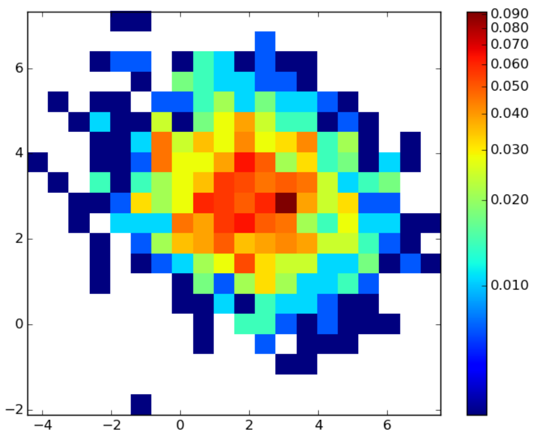

Figure 1: 2D-histogram for 20x20 bins, 1000 samples

plot shows approximately the desired distribution, because the "color density" is maximal around the given μ value, and it decreases exponentially. Furthermore, the standard deviation of the horizontal axis is bigger than the vertical one, and they are negatively correlated, as stated in the Σ matrix.

c) Estimate the mean and covariance matrix from the data using their Maximum Likelihood estimates. How close are the estimated parameters to the real ones? Use 2, 20, 200 data points for your estimate

Run the attached Python2 script to generate similar results:

estimated cov(x, y) for 1000 samples: -0.345968387603

estimated mean of x for 2 samples: 4.82419813156
estimated mean of y for 2 samples: 2.10856773405
estimated variance of x for 2 samples: 0.518180558768
estimated variance of y for2 samples: 0.00268184920367
estimated cov(x, y) for 2 samples: 0.0745568808063

estimated mean of x for 20 samples: 1.87527441792
estimated mean of y for 20 samples: 3.26096396479
estimated variance of x for 20 samples: 2.04668336757
estimated variance of y for20 samples: 2.05992159961
estimated cov(x, y) for 20 samples: -0.0237578274471

estimated mean of x for 200 samples: 2.18152415943
estimated mean of y for 200 samples: 3.17357837903
estimated variance of x for 200 samples: 3.45526621455
estimated variance of y for200 samples: 1.8364091909
estimated cov(x, y) for 200 samples: -0.405670949001

Plot the likelihood as a function of two parameters of your choice, keeping all other parameters constant

The likelihood function for a given set of one-dimensional points $x_1, ..., x_N$ is:

$$\mathcal{L}(\mu, \sigma^2, x_1, ..., x_N) = (2\pi\sigma^2)^{-\frac{n}{2}} exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\{(x_i-\mu)^2\}\right) \tag{8}$$

Since it returns a scalar, it can also easily be expanded to multidimensional points by adding the likelihood of each dimension. Since likelihood is mostly meaningful based on a sample set, it makes sense to calculate a fixed set of points, and take the expected value and variance as variables:

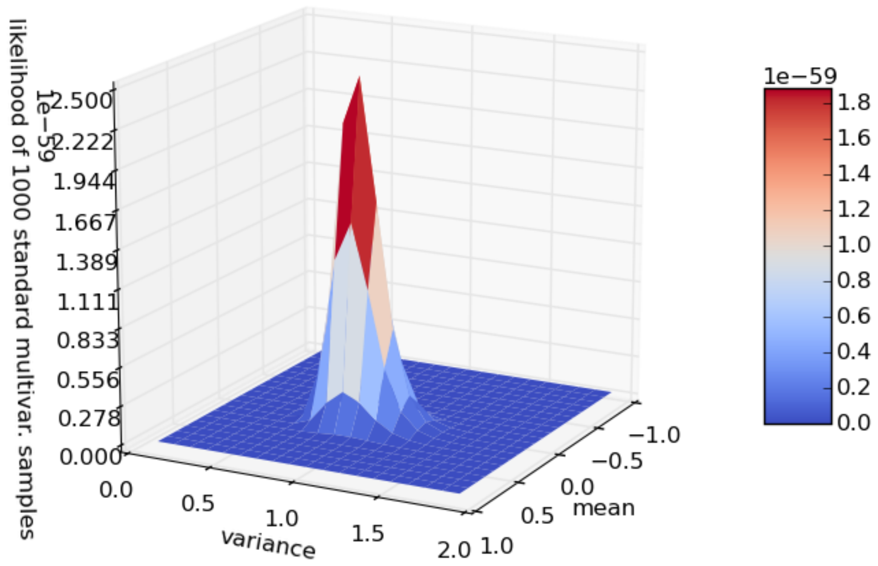

Figure 2: Likelihood of 1000 standard,multivariate, normal distributed samples