Statistical Methods of Machine Learning
Assignment 1

Martin Grunbaum
Martin Jorgensen
Thomas Barnholdt

February 14, 2014

I.1.1.1

Given

$$a = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} b = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$$

Then $a^T b = 9$

I.1.1.2

The *l2-norm* or *Euclidean norm* $||a|| = \sqrt{1^2 + 2^2 + 2^2} = 3$

I.1.1.3

The outer product

$$ab^T = \begin{bmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 6 & 4 & 2 \end{bmatrix}$$

I.1.1.4

The inverse matrix of M is

$$M^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

I.1.1.5

The matrix-vector product $Ma = \begin{pmatrix} 1 \\ 8 \\ 4 \end{pmatrix}$

I.1.1.6

$$A = ab^T = \begin{bmatrix} 3 & 6 & 6 \\ 2 & 4 & 4 \\ 1 & 2 & 2 \end{bmatrix}$$

I.1.1.7

The rank of $A = 1$, because the rows are linearly dependent. We can verify this by observing that the third row can produce the first and second rows with a multiple, e.g. the first row (3 6 6) is the same as the third row (1 2 2) x 3.

I.1.1.8

As A is not full rank, it is not invertible.

I.1.2.1

The derivative of $f(w) = (wx+b)^2$ with respect to w is

$$(wx+b)^2 = w^2x^2 + 2wxb + b^2 = 2x^2w + 2xb = 2x(xw+b)$$

I.1.2.2

In general

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{(g(x))^2}$$

Therefore, differentiating for w we get:

$$\begin{aligned}
f(x) &= 1 \\
f'(x) &= 0 \\
g(x) &= (wx+b)^2 \\
g'(x) &= 2x(wx+b) \\
\left(\frac{f}{g}\right)'(w) &= \frac{0 \cdot (wx+b)^2 - 1 \cdot 2x(wx+b)}{((wx+b)^2)^2} \\
&= \frac{-1 \cdot 2x(wx+b)}{(wx+b)^4} \\
&= \frac{-2x}{(wx+b)^3}
\end{aligned}$$

I.1.2.3

In general

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

Therefore, differentiating for x we get:

$$\begin{aligned}
f(x) &= x \\
g(x) &= e^x \\
(f \cdot g)'(x) &= 1e^x + xe^x
\end{aligned}$$

I.2.1

The plots with gaussian distributions for (sigma, mu) pairs (-1,1), (0,2) and (2,3) can be seen in Figure 1. The code for generating the plots can be found in `unigauss_run.m`, and the code for our gaussian distribution function can be found in `unigauss.m`.

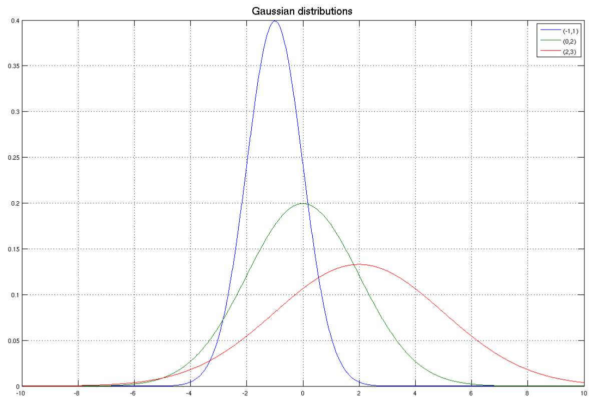

Figure 1: Gaussian distributions plotted with different values for (sigma, mu).

I.2.2

Source code is available in `multigauss.m` and `multigauss_run.m`.

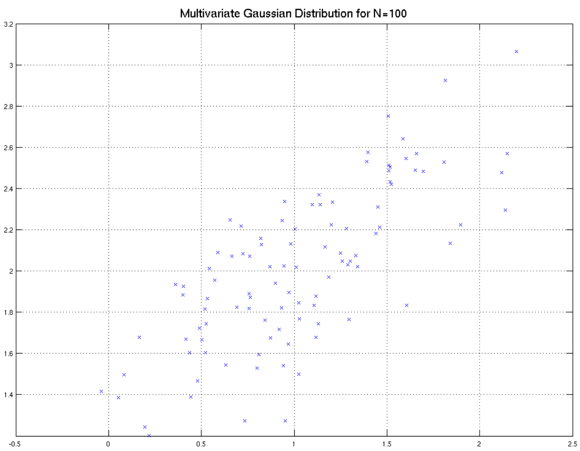

Figure 2: 100 points drawn from a 2-dimensional Multivariate gaussian distribution.

I.2.3

The l2 norm of x is

$$mean = \begin{pmatrix} 1 & 2 \end{pmatrix}^T$$
$$\mu = \begin{pmatrix} 1.0006 & 1.9834 \end{pmatrix}^T$$
$$||x|| = l2(mean - \mu) = 0.0366$$

where $l2()$ is a function that calculates the *Euclidean norm* or l2 norm of the vector $mean - \mu$.

Figure 3 plots the points drawn along with a red circle for the calculated mean and a green circle for μ. There is a difference between the two because the mean is calculated based on the generated data drawn from the multivariate gaussian distribution at random. If we had a number of points approaching infinite, the difference would approach 0.

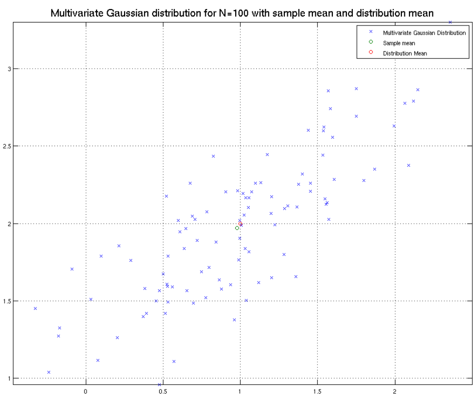

Figure 3: 100 points drawn from a 2-dimensional Multivariate gaussian distribution, plotted with the mean of the distribution and the value of μ.

I.2.4

The covariance matrix is full rank 2 and thus has two eigenvectors and eigenvalues. Each eigenvector represents a principal component (or linearly uncorrelated variable), and each eigenvalue a scalar representing the variance. Intuitively, the eigenvectors form a scaled and translated coordinate system centered at the mean of the multivariate Gaussian distribution (μ). If an eigenvalue is 0, the dimensionality is reduced by one. The larger of the two eigenvector/value pairs represents the direction where the ellipsis is widest.

The covariance matrix we calculated can be found in Eq 1. Figure 4 shows a plot of the Multivariate gaussian distribution, plotted with the mean, μ and the two eigenvectors centered in the distribution μ. Figure 5 shows a plot of the 3 rotated distributions along with the distribution rotated to match the largest eigenvector along the x-axis. The angle needed for this was -37.2564^o in our case.

$$
\begin{aligned}
\Sigma_{ML} &= \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T \\
&= \begin{pmatrix} 0.3239 & 0.2093 \\ 0.2093 & 0.2080 \end{pmatrix}
\end{aligned} \tag{1}
$$

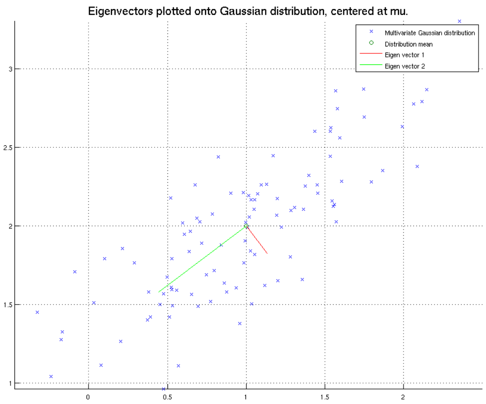

Figure 4: 100 points drawn from a 2-dimensional Multivariate gaussian distribution, plotted with the mean of the distribution, the value of μ and the two eigenvectors centered in the distribution μ.

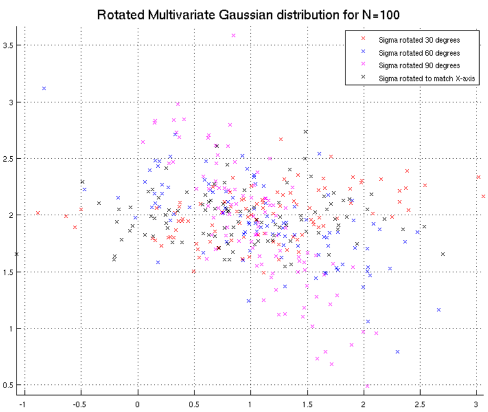

Figure 5: 100 points drawn from a 2-dimensional Multivariate gaussian distribution, rotated at 30, 60 and 90 degrees and lastly also aligned along the x-axis, all distributions in their own color.

I.3

We have μ, $\overline{x}$ and the covariance Σ given by:

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_c \end{pmatrix} \tag{2}$$

$$\overline{x} = \begin{pmatrix} x_a \\ x_b \\ x_c \end{pmatrix} \tag{3}$$

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{bmatrix} \tag{4}$$

We partition our μ, $\overline{x}$ and Σ as follows:

$$\mu = \begin{pmatrix} \mu_d \\ \mu_c \end{pmatrix}, \text{where } \mu_d = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \tag{5}$$

$$\overline{x} = \begin{pmatrix} x_d \\ x_c \end{pmatrix}, \text{where } x_d = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \tag{6}$$

$$\Sigma = \begin{bmatrix} \Sigma_{aad} & \Sigma_{abd} \\ \Sigma_{bad} & \Sigma_{bbd} \end{bmatrix} \tag{7}$$

$$\text{where } \Sigma_{aad} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \tag{8}$$

$$\text{and } \Sigma_{abd} = \begin{bmatrix} \Sigma_{ac} \\ \Sigma_{bc} \end{bmatrix} \tag{9}$$

$$\text{and } \Sigma_{bad} = \begin{bmatrix} \Sigma_{ca} & \Sigma_{cb} \end{bmatrix} \tag{10}$$

$$\text{and } \Sigma_{bbd} = \begin{bmatrix} \Sigma_{cc} \end{bmatrix} \tag{11}$$

We also define a partitioned precision matrix (inverse of the covariance matrix), as:

$$\Lambda \equiv \Sigma^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \tag{12}$$

We wish to now discover $p(\overline{x}_d|\overline{x}_c)$, by considering the quadratic formin the exponent of the Gaussian distribution:

$$-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \tag{13}$$