

Derivation of LDA log likelihood ratio one-to-one classifier

Luuk Spreuwers

The common expression for the Likelihood Ratio classifier using LDA assumes that the reference class mean is available. In biometrics, this is often not the case and only a single sample of the reference class is available. In this paper expressions are derived for biometric comparison between single reference and test samples and M test samples against N reference samples. *Abstract—*

I. LOG LIKELIHOOD RATIO CLASSIFIER WITH 1 ENROLMENT

Consider a biometric classification problem, where we have to compare a probe sample \mathbf{x} to a reference sample \mathbf{y} to determine if \mathbf{x} is of the same class as \mathbf{y} . This is a very typical biometric classification problem, where we have only a single reference sample for enrolment, e.g. for face recognition a good quality frontal photograph or for fingerprint recognition a single high quality fingerprint.

Ideally, we would want to design a classifier that determines $p(c|\mathbf{x})$, i.e. the probability on class c given the sample \mathbf{x} . However, since we only have the reference sample \mathbf{y} to represent class c , we can only determine $p(\text{same class}|\mathbf{x}, \mathbf{y})$, i.e. the probability that the two samples are of the same class given the two samples.

Using Bayes rule, we can write:

$$p(\text{same class}|\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y}|\text{same class})p(\text{same class})}{p(\mathbf{x}, \mathbf{y})} \quad (1)$$

Since often we have no means to determine the prior probability on the occurrence that both samples are of the same class, the best we can do is determine the so-called likelihood ratio of the the occurrence same class:

$$LR(\text{same class}|\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y}|\text{same class})}{p(\mathbf{x}, \mathbf{y})} \quad (2)$$

The probability $p(\mathbf{x}, \mathbf{y}|\text{same class})$ can be expanded as a summation of all occurrences of the samples for the individual classes c_i weighted with the probability on the classes $p(c_i)$:

$$p(\mathbf{x}, \mathbf{y}|\text{same class}) = \sum_i p(\mathbf{x}, \mathbf{y}|c_i)p(c_i) \quad (3)$$

We may assume that the samples \mathbf{x} and \mathbf{y} are independent, so their combined probability is the product of the individual probabilities:

Luuk Spreuwers is with the Chair of Biometric Pattern Recognition, SCS Group, Faculty of EEMCS, University of Twente, Netherlands e-mail: L.j.spreuwers@utwente.nl

Manuscript received 10 November, 2014

$$LR(\text{same class}|\mathbf{x}, \mathbf{y}) = \frac{\sum_i p(\mathbf{x}|c_i)p(\mathbf{y}|c_i)p(c_i)}{p(\mathbf{x})p(\mathbf{y})} \quad (4)$$

In order to be able to design the likelihood ratio classifier, we make the following assumptions:

- The conditional probability density functions $p(\mathbf{x}|c_i)$ are normal with mean μ_i and covariance \mathbf{C}_i
- All classes have different mean, but the same covariance, i.e. $\mathbf{C}_i = \mathbf{C}_w$. This covariance is called the *within class* variance. The variance within a class is e.g. caused by expression and illumination variations in face recognition and assumed to have the same impact on the appearance for every face.
- There is an infinite number of classes c_i , the means of which are normally distributed with mean μ_b and covariance \mathbf{C}_b where the subscript b denotes *between class*.

We can now rewrite equation 4 by substituting the summation by an integral (there is an infinite number of classes) and writing $p_w(\mathbf{x}|\mu)$ for the conditional probability on a sample \mathbf{x} given it is of a class with mean μ . The subscript w means *within class*.

$$LR(\text{same class}|\mathbf{x}, \mathbf{y}) = \frac{\int p_w(\mathbf{x}|\mu)p_w(\mathbf{y}|\mu)p_b(\mu)d\mu}{p_t(\mathbf{x})p_t(\mathbf{y})} \quad (5)$$

The subscript t in the denominator denotes *total*, referring to the total distribution of a sample from any class:

$$p(\mathbf{x}) = p_t(\mathbf{x}) = \sum_i p(\mathbf{x}|c_i)p(c_i) = \int p_w(\mathbf{x}|\mu)p(\mu)d\mu \quad (6)$$

The total mean is denoted μ_t .

II. WHITENING AND DECORRELATING DATA

Now we note that for a translation $\mathbf{x}' = \mathbf{x} + \mathbf{t}$, $\mathbf{y}' = \mathbf{y} + \mathbf{t}$, $\mu' = \mu + \mathbf{t}$, the likelihood ratio expression in equation 5 does not change. All covariances remain the same.

$$LR(\text{same class}|\mathbf{x}, \mathbf{y}) = LR(\text{same class}|\mathbf{x}', \mathbf{y}') \quad (7)$$

Next we note that if we rotate the data with a rotation matrix \mathbf{R} , with $\mathbf{x}' = \mathbf{R}(\mathbf{x})$, $\mathbf{y}' = \mathbf{R}(\mathbf{y})$ and $\mu' = \mathbf{R}(\mu)$, again basically nothing changes with respect to equation 5:

$$LR(\text{same class}|\mathbf{x}, \mathbf{y}) = \frac{\int p_w(\mathbf{x}'|\mu')p_w(\mathbf{y}'|\mu')p_b(\mu')d\mu'}{p_t(\mathbf{x}')p_t(\mathbf{y}')} \quad (8)$$

The covariance matrices after rotation become: $\mathbf{C}'_w = \mathbf{R}^T \mathbf{C}_w \mathbf{R}$ for the within covariance, $\mathbf{C}'_b = \mathbf{R}^T \mathbf{C}_b \mathbf{R}$ for the between covariance and $\mathbf{C}'_t = \mathbf{R}^T \mathbf{C}_t \mathbf{R}$ for the total covariance. The shape of the probability density functions does not change, however.

For a scaling \mathbf{S} of normally distributed data, with $\mathbf{x}' = \mathbf{S}(\mathbf{x})$, with \mathbf{S} a diagonal matrix, the following holds:

$$p_x(\mathbf{x}') = |\mathbf{S}|^{\frac{1}{2}} \mathcal{N}(\mathbf{S}(\mu_x), \mathbf{S} \mathbf{C}_x \mathbf{S}^T) \quad (9)$$

and

$$p_{x'}(\mathbf{x}') = \mathcal{N}(\mathbf{S}(\mu_x), \mathbf{S} \mathbf{C}_x \mathbf{S}^T) \quad (10)$$

Therefore for a transformation \mathbf{T} that translates, rotates and scales the data, the probability density of the transformed data becomes:

$$p_x(\mathbf{x}') = |\mathbf{T}|^{\frac{1}{2}} \mathcal{N}(\mathbf{T}(\mu_x), \mathbf{T} \mathbf{C}_x \mathbf{T}^T) \quad (11)$$

and

$$p_{x'}(\mathbf{x}') = \mathcal{N}(\mathbf{T}(\mu_x), \mathbf{T} \mathbf{C}_x \mathbf{T}^T) \quad (12)$$

The within, between and total density functions of the transformed data then become:

$$\begin{aligned} p_b(\mu') &= \mathcal{N}(\mathbf{T}(\mu_b), \mathbf{T} \mathbf{C}_b \mathbf{T}^T) \\ p_w(\mathbf{x}' | \mu') &= \mathcal{N}(\mathbf{T}(\mu), \mathbf{T} \mathbf{C}_w \mathbf{T}^T) \\ p_t(\mathbf{x}') &= \mathcal{N}(\mathbf{T}(\mu_t), \mathbf{T} \mathbf{C}_t \mathbf{T}^T) \end{aligned} \quad (13)$$

The likelihood ratio can be written as a function of the transformed data:

$$LR(\text{same class} | \mathbf{x}, \mathbf{y}) = \frac{\int_{\mu'} p_w(\mathbf{x}' | \mu') p_w(\mathbf{y}' | \mu') p(\mu') d\mu'}{p_t(\mathbf{x}') p_t(\mathbf{y}')} \quad (14)$$

because the terms $|\mathbf{T}|^{\frac{1}{2}}$ in the nominator and denominator cancel out.

If we choose \mathbf{T} such that it translates the data by μ_t and simultaneously whitens the total distribution and decorrelates the within class distribution, then:

$$\begin{aligned} p_t(\mathbf{x}') &= \mathcal{N}(0, \mathbf{I}) \\ p_w(\mathbf{x}' | \mu') &= \mathcal{N}(\mathbf{T}(\mu), \Sigma_w) \end{aligned} \quad (15)$$

With Σ_w a diagonal matrix and $\mathbf{T}^T \mathbf{C}_t \mathbf{T} = \mathbf{I}$ and $\mathbf{T}^T \mathbf{C}_w \mathbf{T} = \Sigma_w$. The transformation \mathbf{T} can be obtained by singular value decomposition (SVD) from covariances estimated from data. This will be described later.

III. LIKELIHOOD RATIO FOR WHITENED/DECORRELATED DATA

Now for simplified notation, we will simply assume that the original data are transformed such that the total mean is zero, the total covariance matrix is the identity matrix and the within covariance matrix is a diagonal matrix and we will therefore drop the primed notation and simply write \mathbf{x} and μ etc. instead of the primed versions.

The first important step to continue is to determine the background distribution of the transformed data. From the definition of conditional probability, we have:

$$p_t(\mathbf{x}) = \int_{\mu} p_w(\mathbf{x} | \mu) p(\mu) d\mu \quad (16)$$

Filling in the known densities, we obtain for the right hand term:

$$\int_{\mu} \frac{e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma_w^{-1}(\mathbf{x}-\mu)}}{(2\pi)^{\frac{m}{2}} |\Sigma_w|^{\frac{1}{2}}} \cdot \frac{e^{-\frac{1}{2}(\mu-\mu_b)^T \mathbf{C}_b^{-1}(\mu-\mu_b)}}{(2\pi)^{\frac{m}{2}} |\mathbf{C}_b|^{\frac{1}{2}}} d\mu \quad (17)$$

Where m is the dimensionality of the data. Combining the two terms in the integral and considering just the exponential, we can write:

$$\begin{aligned} -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma_w^{-1}(\mathbf{x}-\mu) - \frac{1}{2}(\mu-\mu_b)^T \mathbf{C}_b^{-1}(\mu-\mu_b) &= \\ = -\frac{1}{2}(\mu-\alpha)^T \mathbf{C}_\alpha^{-1}(\mu-\alpha) - \frac{1}{2}R \end{aligned} \quad (18)$$

After some manipulation, we find:

$$\begin{aligned} \mathbf{C}_\alpha^{-1} &= \Sigma_w^{-1} + \mathbf{C}_b^{-1} \\ \alpha &= \mathbf{C}_\alpha [\Sigma_w^{-1} \mathbf{x} + \mathbf{C}_b^{-1} \mu_b] \\ R &= \mathbf{x}^T \Sigma_w^{-1} \mathbf{x} + \mu_b^T \mathbf{C}_b^{-1} \mu_b - \alpha^T \mathbf{C}_\alpha^{-1} \alpha \end{aligned} \quad (19)$$

Since R does not depend on μ , it can be taken outside the integral. Combining equations 17 and 18, we get:

$$\frac{1}{(2\pi)^m |\Sigma_w|^{\frac{1}{2}} |\mathbf{C}_b|^{\frac{1}{2}}} e^{-\frac{1}{2}R} \int_{\mu} e^{-\frac{1}{2}(\mu-\alpha)^T \mathbf{C}_\alpha^{-1}(\mu-\alpha)} d\mu \quad (20)$$

The integral is now just the integral over a Gaussian, resulting in $(2\pi)^{\frac{m}{2}} |\mathbf{C}_\alpha|^{\frac{1}{2}}$, so we get:

$$p_t(\mathbf{x}) = \frac{(2\pi)^{\frac{m}{2}} |\mathbf{C}_\alpha|^{\frac{1}{2}}}{(2\pi)^m |\Sigma_w|^{\frac{1}{2}} |\mathbf{C}_b|^{\frac{1}{2}}} e^{-\frac{1}{2}R} \quad (21)$$

Since the $p_t(\mathbf{x})$ is normally distributed with zero mean and the identity matrix as a covariance matrix, it immediately follows from equation 19 that μ_b must be zero as well and \mathbf{C}_α and hence \mathbf{C}_b must be diagonal. It is now easy to see from equations 19 and 21 that $\mathbf{C}_\alpha = \Sigma_w \mathbf{C}_b$ and finally:

$$\mathbf{C}_b + \Sigma_w = \mathbf{I} \quad (22)$$

Thus we can conclude that if the samples are distributed normally with mean zero and the identity matrix as a covariance matrix and the within class distribution is normal with mean μ is uncorrelated, i.e. its covariance matrix Σ_w is a diagonal matrix, then the between distribution of μ is normal with mean zero and uncorrelated with covariance $\mathbf{C}_b = \mathbf{I} - \Sigma_w$. Since \mathbf{C}_b is diagonal, we will write Σ_b from now on.

We now return to the expression for the likelihood ratio and use the obtained knowledge:

$$LR(\text{same class} | \mathbf{x}, \mathbf{y}) = \frac{\int_{\mu} p_w(\mathbf{x} | \mu) p_w(\mathbf{y} | \mu) p_b(\mu) d\mu}{p_t(\mathbf{x}) p_t(\mathbf{y})} =$$

$$\begin{aligned}
& \frac{\int_{\mu} e^{-\frac{1}{2}[(\mathbf{x}-\mu)^T \Sigma_w^{-1}(\mathbf{x}-\mu) + (\mathbf{y}-\mu)^T \Sigma_w^{-1}(\mathbf{y}-\mu) + \mu^T \Sigma_b^{-1} \mu]} d\mu}{(2\pi)^{\frac{3m}{2}} |\Sigma_w| |\Sigma_b|^{\frac{1}{2}}} = \\
& \frac{1}{(2\pi)^m |\mathbf{I}|} e^{-\frac{1}{2}[\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}]} \\
& \frac{\int_{\mu} e^{-\frac{1}{2}[(\mathbf{x}-\mu)^T \Sigma_w^{-1}(\mathbf{x}-\mu) + (\mathbf{y}-\mu)^T \Sigma_w^{-1}(\mathbf{y}-\mu) + \mu^T \Sigma_b^{-1} \mu]} d\mu}{(2\pi)^{\frac{m}{2}} |\Sigma_w| |\Sigma_b|^{\frac{1}{2}} e^{-\frac{1}{2}[\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}]}} = \\
& \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_w| |\Sigma_b|^{\frac{1}{2}}} \frac{\int_{\mu} e^{-\frac{1}{2}[(\mu-\alpha)^T \Sigma_{\alpha}^{-1}(\mu-\alpha) + R]} d\mu}{e^{-\frac{1}{2}[\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}]}} \quad (23)
\end{aligned}$$

After some manipulation, we obtain:

$$\begin{aligned}
\Sigma_{\alpha}^{-1} &= 2\Sigma_w^{-1} + \Sigma_b^{-1} \\
\alpha &= \Sigma_{\alpha} [\Sigma_w^{-1} \mathbf{x} + \Sigma_w^{-1} \mathbf{y}] \\
R &= \mathbf{x}^T \Sigma_w^{-1} \mathbf{x} + \mathbf{y}^T \Sigma_w^{-1} \mathbf{y} - \alpha^T \Sigma_{\alpha}^{-1} \alpha \quad (24)
\end{aligned}$$

Since R is independent of μ , it can be brought before the integral. The integral itself is then the integral over a complete Gaussian and reduces to $(2\pi)^{\frac{m}{2}} |\Sigma_{\alpha}|^{\frac{1}{2}}$

The likelihood ratio thus becomes, dropping the $(2\pi)^{\frac{m}{2}}$ terms in nominator and denominator:

$$\frac{|\Sigma_{\alpha}|^{\frac{1}{2}} e^{-\frac{1}{2}R}}{|\Sigma_w| |\Sigma_b|^{\frac{1}{2}} e^{-\frac{1}{2}[\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}]}} \quad (25)$$

Substituting α and Σ_{α} into R in equation 24 gives:

$$\begin{aligned}
R &= \mathbf{x}^T \Sigma_w^{-1} \mathbf{x} + \mathbf{y}^T \Sigma_w^{-1} \mathbf{y} - \\
& (\mathbf{x} + \mathbf{y})^T \Sigma_w^{-1} [2\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} (\mathbf{x} + \mathbf{y}) \quad (26)
\end{aligned}$$

Since all matrices are diagonal matrices, all dimensions are decorrelated and we can treat each dimension separately. If we write σ_{wi}^2 for element (i, i) of the diagonal matrix Σ_w and σ_{bi}^2 for element (i, i) of the diagonal matrix Σ_b we obtain:

$$R = \sum_i \frac{x_i^2 + y_i^2}{\sigma_{wi}^2} - (x_i + y_i)^2 \frac{\sigma_{bi}^2}{\sigma_{wi}^2 (2\sigma_{bi}^2 + \sigma_{wi}^2)} \quad (27)$$

Substitution into equation 25 and combining nominator and denominator, we get the following expression for the likelihood ratio:

$$\frac{|\Sigma_{\alpha}|^{\frac{1}{2}}}{|\Sigma_w| |\Sigma_b|^{\frac{1}{2}}} e^{-\frac{1}{2} \left[\sum_i \frac{x_i^2 + y_i^2}{\sigma_{wi}^2} - (x_i + y_i)^2 \frac{\sigma_{bi}^2}{\sigma_{wi}^2 (2\sigma_{bi}^2 + \sigma_{wi}^2)} - x_i^2 - y_i^2 \right]} \quad (28)$$

For some constants Q , ρ_i and d_i we can write for the expression inside the summation:

$$(x_i - \rho_i y_i)^2 d_i + Q - x_i^2 - y_i^2 \quad (29)$$

Using the property $\sigma_{wi}^2 + \sigma_{bi}^2 = 1$, which follows from $\Sigma_w + \Sigma_b = \mathbf{I}$ and some manipulation it follows that:

$$\begin{aligned}
d_i &= \frac{1}{\sigma_{wi}^2 (2\sigma_{bi}^2 + \sigma_{wi}^2)} = \frac{1}{\sigma_{wi}^2 (2 - \sigma_{wi}^2)} \\
\rho_i &= \sigma_{bi}^2 = 1 - \sigma_{wi}^2 \\
Q &= y_i^2 \quad (30)
\end{aligned}$$

Substitution into equation 28 and going back to vector matrix notation and introducing the diagonal matrices \mathbf{P} and \mathbf{D} with the elements ρ_i resp d_i on their diagonals, we obtain the final simple expression for the likelihood ratio:

$$\begin{aligned}
LR(\text{same class} | \mathbf{x}, \mathbf{y}) &= \\
& \frac{[2\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1}}{|\Sigma_w| |\Sigma_b|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\mathbf{x} - \mathbf{P}\mathbf{y})^T \mathbf{D}(\mathbf{x} - \mathbf{P}\mathbf{y}) - \mathbf{x}^T \mathbf{x}]} \quad (31)
\end{aligned}$$

If we take the logarithm and ignore the constants that only offset the results, we obtain a very simple expression for a log likelihood ratio based score L :

$$L(\text{same class} | \mathbf{x}, \mathbf{y}) = -(\mathbf{x} - \mathbf{P}\mathbf{y})^T \mathbf{D}(\mathbf{x} - \mathbf{P}\mathbf{y}) + \mathbf{x}^T \mathbf{x} \quad (32)$$

Or we could use equation 26 resulting in:

$$\begin{aligned}
L(\text{same class} | \mathbf{x}, \mathbf{y}) &= -\mathbf{x}^T \Sigma_w^{-1} \mathbf{x} - \mathbf{y}^T \Sigma_w^{-1} \mathbf{y} + \\
& + (\mathbf{x} + \mathbf{y})^T \Gamma (\mathbf{x} + \mathbf{y}) + \mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y} = \\
& = \mathbf{x}^T \Lambda \mathbf{x} + \mathbf{y}^T \Lambda \mathbf{y} + (\mathbf{x} + \mathbf{y})^T \Gamma (\mathbf{x} + \mathbf{y}) \quad (33)
\end{aligned}$$

With Γ and Λ defined as:

$$\begin{aligned}
\Gamma &= \Sigma_w^{-1} [2\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \\
\Lambda &= \mathbf{I} - \Sigma_w^{-1} \quad (34)
\end{aligned}$$

IV. ESTIMATING STATISTICS FROM TRAINING DATA

For practical use of the score on real biometric data, first the matrices Σ_w and Σ_b have to be estimated from training data. Furthermore, a transformation \mathbf{T} was introduced that translates the data by μ_t and simultaneously whitens the total distribution and decorrelates the within class distribution. The transformation and the mean of the data μ_t also have to be determined from the training data.

Suppose we have n sample vectors \mathbf{m}_i of dimension d , ordered in a $n \times d$ matrix \mathbf{M} , then unbiased estimators for the total mean and covariance are:

$$\hat{\mu}_t = \frac{1}{n} \sum_1^n \mathbf{m}_i \quad (35)$$

If the zero mean vectors z_i are defined by:

$$\mathbf{z}_i = \mathbf{m}_i - \hat{\mu}_t \quad (36)$$

and the zero mean vectors \mathbf{z}_i are again ordered in a $n \times d$ matrix \mathbf{Z} , then:

$$\hat{\mathbf{C}}_t = \frac{1}{n-1} \mathbf{Z} \mathbf{Z}^T \quad (37)$$

Using the singular value decomposition (SVD) of \mathbf{Z} , we can write:

$$\mathbf{Z} = \mathbf{U}_t \mathbf{S}_t \mathbf{V}_t^T \quad (38)$$

Where \mathbf{U}_t and \mathbf{V}_t are the left and right singular vectors (i.e. the eigenvectors of $\mathbf{Z} \mathbf{Z}^T$ resp. $\mathbf{Z}^T \mathbf{Z}$) and \mathbf{S}_t is a diagonal matrix with the singular values of \mathbf{Z} , which are the square roots of the eigenvalues of $\mathbf{Z} \mathbf{Z}^T$. An important property of \mathbf{U}_t and \mathbf{V}_t are that they are unitary matrices, i.e. $\mathbf{U}_t \mathbf{U}_t^T = \mathbf{I}$ and $\mathbf{V}_t \mathbf{V}_t^T = \mathbf{I}$.

We can now write for \mathbf{C}_t :

$$\hat{\mathbf{C}}_t = \frac{1}{n-1} \mathbf{Z}\mathbf{Z}^T = \mathbf{U}_t \frac{\mathbf{S}_t \mathbf{S}_t}{n-1} \mathbf{U}_t^T = \mathbf{U}_t \hat{\Sigma}_t \mathbf{U}_t^T \quad (39)$$

The transformation that whitens the total distribution is:

$$\mathbf{T}_1 = \sqrt{n-1} \mathbf{S}_t^{-1} \mathbf{U}_t^T \quad (40)$$

because:

$$\mathbf{T}_1 \hat{\mathbf{C}}_t \mathbf{T}_1^T = \mathbf{I} \quad (41)$$

Next we transform all sample vectors by subtracting the total mean μ_t and multiplying with \mathbf{T}_1 .

$$\mathbf{m}'_i = \mathbf{T}_1 (\mathbf{m}_i - \hat{\mu}_t) \quad (42)$$

To estimate \mathbf{C}_w , of the transformed data, we assume that the sample vectors are of various classes c and for each class there are n_c samples available. An unbiased estimator for the mean of each class is obtained by summing over the sample vectors of the specific class and dividing by the number of samples of the specific class:

$$\hat{\mu}_{ci} = \frac{1}{n_c} \sum_{i: \mathbf{m}_i \in c} \mathbf{m}'_i \quad (43)$$

If we write the class zero mean vectors \mathbf{z}_{ci} as:

$$\mathbf{z}_{ci} = \mathbf{m}'_i - \hat{\mu}_{ci} \quad (44)$$

Where $\hat{\mu}_{ci}$ is the estimate of the class mean of sample vector \mathbf{m}'_i . If we order the class zero mean vectors \mathbf{z}_{ci} into an $n \times d$ matrix \mathbf{Z}_c , we can estimate \mathbf{C}_w by:

$$\hat{\mathbf{C}}_w = \frac{1}{n-1} \mathbf{Z}_c \mathbf{Z}_c^T \quad (45)$$

Using SVD, we can write

$$\hat{\mathbf{C}}_w = \frac{1}{n-1} \mathbf{Z}_c \mathbf{Z}_c^T = \mathbf{U}_w \frac{\mathbf{S}_w \mathbf{S}_w}{n-1} \mathbf{U}_w^T = \mathbf{U}_w \hat{\Sigma}_w \mathbf{U}_w^T \quad (46)$$

And we find the transformation that decorrelates the within distribution as:

$$\mathbf{T}_2 = \mathbf{U}_w^T \quad (47)$$

The total transformation \mathbf{T} is the product of the two transformations:

$$\mathbf{T} = \mathbf{T}_2 \mathbf{T}_1 = \mathbf{U}_w^T \sqrt{n-1} \mathbf{S}_t^{-1} \mathbf{U}_t^T \quad (48)$$

The second transformation does not alter the total distribution, because it is white anyway and \mathbf{T}_2 is a rotation:

$$\mathbf{T}_2 \mathbf{T}_1 \hat{\mathbf{C}}_t \mathbf{T}_1^T \mathbf{T}_2^T = \mathbf{T}_2 \mathbf{I} \mathbf{T}_2^T = \mathbf{U}_w^T \mathbf{I} \mathbf{U}_w = \mathbf{I} \quad (49)$$

We now have all required statistical information: the transformation \mathbf{T} , an estimate for Σ_w and an estimate for the total sample mean μ_t to calculate the log likelihood ratio based score for two samples using equations 32 and 30.

V. DIMENSIONALITY REDUCTION USING PCA AND LDA

Generally there is insufficient data available to obtain a good estimate of the full covariance matrix \mathbf{C}_t .

A solution to this problem is to perform principle component analysis (PCA). This means instead of estimating the full covariance matrix, only the components with the largest variation are estimated. The variation in the different directions is given by the eigenvalues (or singular values).

In practice this means that if the eigenvalues (or singular values) are ordered from large to small, only the first p largest eigenvalues/singular values and corresponding eigenvectors are estimated and the rest are set to 0. Using the SVD, we then obtain an approximation \mathbf{C}_{tp} of \mathbf{C}_t :

$$\mathbf{C}_{tp} \approx \mathbf{U}_{tp} \hat{\Sigma}_{tp} \mathbf{U}_{tp}^T \quad (50)$$

Where \mathbf{U}_{tp} is the $n \times p$ matrix consisting of the first p eigenvectors and Σ_{tp} is the $p \times p$ diagonal matrix of the first p eigen values.

The best discrimination between the classes is obtained by projecting the vectors on the subspace with the l smallest eigenvalues/singular values. This means that we can also perform a dimensionality reduction on the within covariance. Since the within covariance is determined on the transformed data using \mathbf{T}_1 , the number of dimensions already is reduced to p . This means only the smallest l eigenvalues of the p remaining eigenvalues and corresponding eigenvectors are used resulting in \mathbf{U}_{wl} and Σ_{wl} .

This results in a $d \times l$ transformation matrix:

$$\mathbf{T}_{pl} = \mathbf{T}_{2l} \mathbf{T}_{1p} = \mathbf{U}_{wl}^T \sqrt{n-1} \mathbf{S}_{tp}^{-1} \mathbf{U}_{tp}^T \quad (51)$$

VI. MULTIPLE ENROLMENT

The likelihood framework is easily extended to multiple enrolment samples. If we have N enrolment samples $\mathbf{y}_1 \dots \mathbf{y}_N$, then:

$$LR(\text{same class} | \mathbf{x}, \mathbf{y}_1 \dots \mathbf{y}_N) = \frac{p(\mathbf{x}, \mathbf{y}_1 \dots \mathbf{y}_N | \text{same class})}{p(\mathbf{x}) p(\mathbf{y}_1 \dots \mathbf{y}_N | \text{same class})} \quad (52)$$

Assuming normal distributions again and the same within covariance for all classes and a normally class mean distribution again leads to:

$$LR(\text{same class} | \mathbf{x}, \mathbf{y}_1 \dots \mathbf{y}_N) = \frac{\int p_w(\mathbf{x} | \mu) \prod_{i=1}^N p_w(\mathbf{y}_i | \mu) p_b(\mu) d\mu}{p_t(\mathbf{x}) \int \prod_{i=1}^N p_w(\mathbf{y}_i | \mu) p_b(\mu) d\mu} \quad (53)$$

In a similar fashion as equations 23-26, we can rewrite the likelihood ratio as:

$$\frac{1}{|\Sigma_w|^{1/2}} \frac{\int_{\mu} e^{-\frac{1}{2} [(\mu - \alpha)^T \Sigma_{\alpha}^{-1} (\mu - \alpha) + R]} d\mu}{e^{-\frac{1}{2} \mathbf{x}^T \mathbf{x}} \int_{\mu} e^{-\frac{1}{2} [(\mu - \beta)^T \Sigma_{\beta}^{-1} (\mu - \beta) + Q]} d\mu} \quad (54)$$

Where:

$$R = \mathbf{x}^T \Sigma_w^{-1} \mathbf{x} + \sum_{i=1}^N \mathbf{y}_i^T \Sigma_w^{-1} \mathbf{y}_i - (\mathbf{x} + \sum_{i=1}^N \mathbf{y}_i)^T \Gamma_\alpha (\mathbf{x} + \sum_{i=1}^N \mathbf{y}_i)$$

$$Q = \sum_{i=1}^N \mathbf{y}_i^T \Sigma_w^{-1} \mathbf{y}_i - \left(\sum_{i=1}^N \mathbf{y}_i \right)^T \Gamma_\beta \left(\sum_{i=1}^N \mathbf{y}_i \right) \quad (55)$$

With:

$$\Gamma_\alpha = \Sigma_w^{-1} [(N+1)\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \quad (56)$$

$$\Gamma_\beta = \Sigma_w^{-1} [N\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1}$$

This in turn results in an expression for the log likelihood ratio based score (ignoring constants) of:

$$\begin{aligned} L(\text{same class} | \mathbf{x}, \mathbf{y}_1 \dots \mathbf{y}_N) &= \\ &= \mathbf{x}^T \Lambda \mathbf{x} + (\mathbf{x} + \sum_{i=1}^N \mathbf{y}_i)^T \Gamma_\alpha (\mathbf{x} + \sum_{i=1}^N \mathbf{y}_i) - \left(\sum_{i=1}^N \mathbf{y}_i \right)^T \Gamma_\beta \left(\sum_{i=1}^N \mathbf{y}_i \right) \end{aligned} \quad (57)$$

With Λ defined as:

$$\Lambda = \mathbf{I} - \Sigma_w^{-1} \quad (58)$$

A special interesting case is if $N \rightarrow \infty$. In that case the average of all enrolment samples is an exact estimate of the class mean:

$$\sum_{i=1}^N \mathbf{y}_i = N\mu_c \quad (59)$$

Furthermore, Σ_b can be neglected in the expressions for Γ_α and Γ_β which then simplify to $\frac{1}{N+1}\Sigma_w^{-1}$ resp. $\frac{1}{N}\Sigma_w^{-1}$.

Substitution into equation 57 gives:

$$\begin{aligned} L(c|\mathbf{x}) &= \mathbf{x}^T \Lambda \mathbf{x} + (\mathbf{x} + N\mu_c)^T \frac{1}{N+1} \Sigma_w^{-1} (\mathbf{x} + N\mu_c) - \\ &+ (N\mu_c)^T \frac{1}{N} \Sigma_w^{-1} (N\mu_c) = \\ &= -(\mathbf{x} - \mu_c)^T \Sigma_w^{-1} (\mathbf{x} - \mu_c) + \mathbf{x}^T \mathbf{x} \end{aligned} \quad (60)$$

The well known expression for the log likelihood ratio based score for two classes for known class mean.

VII. MULTIPLE PROBE SAMPLES

Another logical extension is to allow more than a single probe sample. If we have M probe samples $\mathbf{x}_1 \dots \mathbf{x}_M$, then:

$$\begin{aligned} LR(\text{same class} | \mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N) &= \\ &= \frac{p(\mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N | \text{same class})}{p(\mathbf{x}_1 \dots \mathbf{x}_M | \text{same class}) p(\mathbf{y}_1 \dots \mathbf{y}_N | \text{same class})} \end{aligned} \quad (61)$$

Assuming normal distributions again and the same within covariance for all classes and a normally class mean distribution again leads to:

$$\begin{aligned} LR(\text{same class} | \mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N) &= \\ &= \frac{\int \prod_{\mu} \prod_{j=1}^M p_w(\mathbf{x}_j | \mu) \prod_{i=1}^N p_w(\mathbf{y}_i | \mu) p_b(\mu) d\mu}{\int \prod_{\mu} \prod_{j=1}^M p_w(\mathbf{x}_j | \mu) p_b(\mu) d\mu \int \prod_{\mu} \prod_{i=1}^N p_w(\mathbf{y}_i | \mu) p_b(\mu) d\mu} \end{aligned} \quad (62)$$

Which results in the general expression for the log likelihood ratio based score for N enrolments and M probes:

$$\begin{aligned} L(\text{same class} | \mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N) &= -M\bar{\mathbf{x}}^T \Gamma_\gamma M\bar{\mathbf{x}} + \\ &+ (M\bar{\mathbf{x}} + N\bar{\mathbf{y}})^T \Gamma_\alpha (M\bar{\mathbf{x}} + N\bar{\mathbf{y}}) - N\bar{\mathbf{y}}^T \Gamma_\beta N\bar{\mathbf{y}} \end{aligned} \quad (63)$$

With $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ the averages of the probe and enrolment samples (or $M\bar{\mathbf{x}}$ and $N\bar{\mathbf{y}}$ the sum of all probe resp. enrolment samples) and:

$$\begin{aligned} \Gamma_\alpha &= \Sigma_w^{-1} [(M+N)\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \\ \Gamma_\beta &= \Sigma_w^{-1} [N\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \\ \Gamma_\gamma &= \Sigma_w^{-1} [M\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \end{aligned} \quad (64)$$

We can also write this in the form of:

$$\begin{aligned} L(\text{same class} | \mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N) &= \\ &= \bar{\mathbf{x}}^T \Gamma_x \bar{\mathbf{x}} + 2\bar{\mathbf{x}}^T \Gamma_{xy} \bar{\mathbf{y}} + \bar{\mathbf{y}}^T \Gamma_y \bar{\mathbf{y}} \end{aligned} \quad (65)$$

Where:

$$\begin{aligned} \Gamma_x &= M^2(\Gamma_\alpha - \Gamma_\gamma) = \frac{M^2}{N} \Sigma_w^{-1} \\ \Gamma_{xy} &= MN\Gamma_\alpha = MN\Sigma_w^{-1} [(M+N)\Sigma_w^{-1} + \Sigma_b^{-1}]^{-1} \Sigma_w^{-1} \\ \Gamma_y &= N^2(\Gamma_\alpha - \Gamma_\beta) = \frac{N^2}{M} \Sigma_w^{-1} \end{aligned} \quad (66)$$

If we consider $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ as a combined vector, this can also be written as:

$$\begin{aligned} L(\text{same class} | \mathbf{x}_1 \dots \mathbf{x}_M, \mathbf{y}_1 \dots \mathbf{y}_N) &= \\ &= \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix}^T \begin{bmatrix} \Gamma_x & \Gamma_{xy} \\ \Gamma_{xy} & \Gamma_y \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix} \end{aligned} \quad (67)$$

REFERENCES

- [1] Raymond Veldhuis and Asker Bazén, "One-to-template and one-to-one verification in the single- and multi-user case", in: *Twenty-sixth Symposium on Information Theory in the Benelux*, Brussels, Belgium, May 19-20, 2005, pages 39-46