Biometric score normalization - problem formulation and solutions

mgr inż. Radosław Białobrzeski

ZBUM seminar, 17.04.2018

Different people - different similarity rates

The degree of similarity between certain people's traits and **everyone else's** differs significantly. Common or subtle features (e.g. facial) of a given subject may alter the recognition quality. Particularly uncommon features may result in the same. This is the case both in human and automated recognition.

On sample level, obviously, the above statement is just as correct. Hence the possible benefit of applying additional score normalization.

The Biometric Menagerie

An analysis of the varying score distributions was done in: Yager and Dunstone 2010. A few distinct score distributions were proposed:

- sheep - high-intra,
- goats - low-intra,
- lambs - high-inter (always),
- wolves - high-inter (when impersonating).

No generative assumptions were made and only empirical evidence was studied, somehow proving the obvious differences in score distributions and fitting them in the Biometric Menagerie framework.

A statistical approach

By adopting a statistical framework, one can transform score distributions and achieve comparable similarity metrics for **every subject/sample**, therefore enhancing overall performance of a biometric system. Three distinct steps are required for statistically correct score normalization:

- knowledge of a generative model and its properties,
- score distributions estimation,
- score distributions transformation.

Factor analysis with normal priors

Let's explore a particular generative model for factor analysis with the $\boldsymbol{h}_n$ identity factor for the n-th subject:

$$\boldsymbol{h}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$
$$\boldsymbol{x}_n | \boldsymbol{h}_n \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{V}\boldsymbol{h}_n, \boldsymbol{\Sigma}).$$

Of course, one can employ additional latent factors - channel, duration, synthesis etc.

PLDA

By not restricting the Σ covariance matrix to be diagonal (assimilating e.g. channel variability), the FA generative model can be used in probabilistic linear discriminant analysis (Prince and Elder 2007), where the log-likelihood of two samples having the same latent identity variable is:

$$s_{log}(\boldsymbol{x}_r, \boldsymbol{x}_t) = -\begin{bmatrix}\boldsymbol{x}_r' & \boldsymbol{x}_t'\end{bmatrix}\begin{bmatrix}\Sigma + \boldsymbol{V}\boldsymbol{V}' & \boldsymbol{V}\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{V}' & \Sigma + \boldsymbol{V}\boldsymbol{V}'\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{x}_r \\ \boldsymbol{x}_t\end{bmatrix} + \\ + \boldsymbol{x}_r'\left[\Sigma + \boldsymbol{V}\boldsymbol{V}'\right]^{-1}\boldsymbol{x}_r + \boldsymbol{x}_t'\left[\Sigma + \boldsymbol{V}\boldsymbol{V}'\right]^{-1}\boldsymbol{x}_t + const.$$

Notice - this formula assumes zero mean, i.e. subtracting an estimated mean from every sample.

Distribution of PLDA scores

The log-likelihood is a function of two random variables and model parameters. By fixing either the latent variable (subject level) or the sample (sample level), one can further investigate the score distribution.

Unfortunately, both subject and sample scores follow a **non-central generalized χ^2 distribution**. There is both no closed-form solution for the PDF of n.c.g. χ^2 and no generalized, parametric normalization procedure. Still, expectations and variances can be derived from the previous formulas to show their variability and perform at least basic standardization.

Deriving the formulas

In order to find the solutions for expected values and variances of PLDA scores, a set of properties for various algebraic entities was needed. The properties were all found in an **extremely** helpful book - "The Matrix Cookbook" (Petersen and Pedersen 2012).

All of the following formulas were verified in extensive Monte-Carlo simulations.

Fixed identity - expectation

Given: $\boldsymbol{\Sigma}, \boldsymbol{V}, \boldsymbol{h}_r$

$$\boldsymbol{x}_r|\boldsymbol{h}_r \sim \mathcal{N}(\boldsymbol{V}\boldsymbol{h}_r; \boldsymbol{\Sigma}) \quad \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma} + \boldsymbol{V}\boldsymbol{V}' \quad \boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{0}; \boldsymbol{\Sigma}_t)$$

$$\boldsymbol{\mu}_{rt} = \begin{bmatrix} \boldsymbol{V}\boldsymbol{h}_r \\ \boldsymbol{0} \end{bmatrix} \quad \boldsymbol{\Sigma}_{rt} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_t \end{bmatrix}$$

$$\boldsymbol{Q}_1 = \begin{bmatrix} \boldsymbol{\Sigma}_t & \boldsymbol{V}\boldsymbol{V}' \\ \boldsymbol{V}\boldsymbol{V}' & \boldsymbol{\Sigma}_t \end{bmatrix} \quad \boldsymbol{Q}_2 = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_t^{-1} \end{bmatrix} \quad \boldsymbol{Q}_3 = \begin{bmatrix} \boldsymbol{\Sigma}_t^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

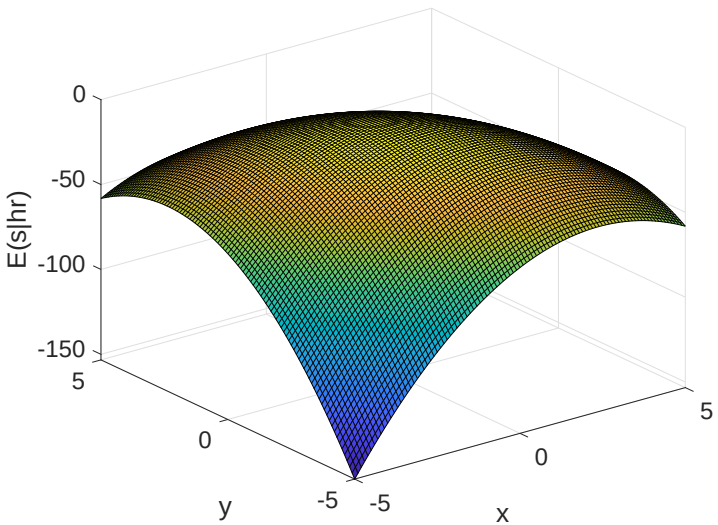
$$\mathbb{E}[s_{log}|\boldsymbol{h}_r] - const. = \mathrm{tr}(\boldsymbol{I}) + \mathrm{tr}(\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}) + (\boldsymbol{V}\boldsymbol{h}_r)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{V}\boldsymbol{h}_r) - \\ - \mathrm{tr}(\boldsymbol{Q}_1^{-1}\boldsymbol{\Sigma}_{rt}) - \boldsymbol{\mu}_{rt}'\boldsymbol{Q}_1^{-1}\boldsymbol{\mu}_{rt}$$

Fixed identity - variance

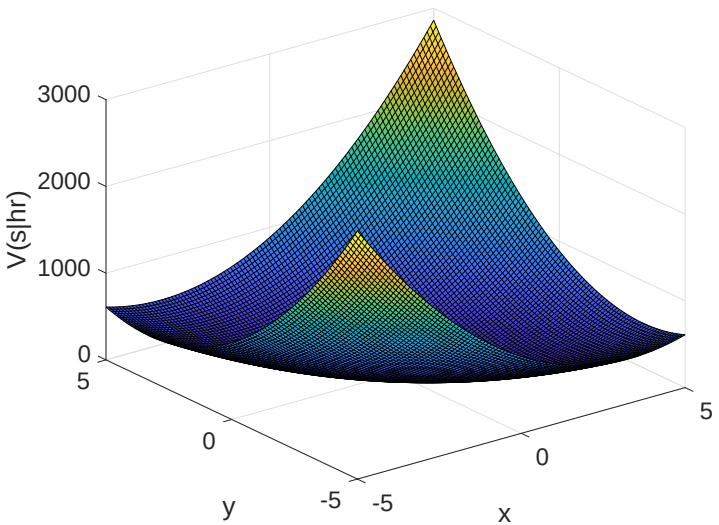
$$W_0 = \Sigma_t^{-1} + (\Sigma_t^{-1})' \quad W_1 = Q_1^{-1} + (Q_1^{-1})' \quad W_2 = Q_2 + Q_2' \quad W_3 = Q_3 + Q_3'$$

$$\begin{aligned}\mathbb{V}[s_{log}|\boldsymbol{h}_r] = {} & \mathrm{tr}(\boldsymbol{Q}_1^{-1}\Sigma_{rt}\boldsymbol{W}_1)\Sigma_{rt} + \boldsymbol{\mu}_{rt}'\boldsymbol{W}_1\Sigma_{rt}\boldsymbol{W}_1\boldsymbol{\mu}_{rt} + \\ & + \mathrm{tr}(\Sigma_t^{-1}\Sigma\boldsymbol{W}_0)\Sigma + (\boldsymbol{V}\boldsymbol{h}_r)'\boldsymbol{W}_0\Sigma\boldsymbol{W}_0(\boldsymbol{V}\boldsymbol{h}_r) + \\ & + \mathrm{tr}(\boldsymbol{W}_0\Sigma_t) - \\ & - 2\big(\mathrm{tr}(\boldsymbol{Q}_1^{-1}\Sigma_{rt}\boldsymbol{W}_2)\Sigma_{rt} + \boldsymbol{\mu}_{rt}'\boldsymbol{W}_1\Sigma_{rt}\boldsymbol{W}_2\boldsymbol{\mu}_{rt}\big) - \\ & - 2\big(\mathrm{tr}(\boldsymbol{Q}_1^{-1}\Sigma_{rt}\boldsymbol{W}_3)\Sigma_{rt} + \boldsymbol{\mu}_{rt}'\boldsymbol{W}_1\Sigma_{rt}\boldsymbol{W}_3\boldsymbol{\mu}_{rt}\big)\end{aligned}$$

Fixed identity - example



(a) Expectations for 2D identity variables and a given model.



(b) Variances for 2D identity variables and a given model.

Fixed sample - expectation

Given: $\Sigma, \boldsymbol{V}, \boldsymbol{x}_r$

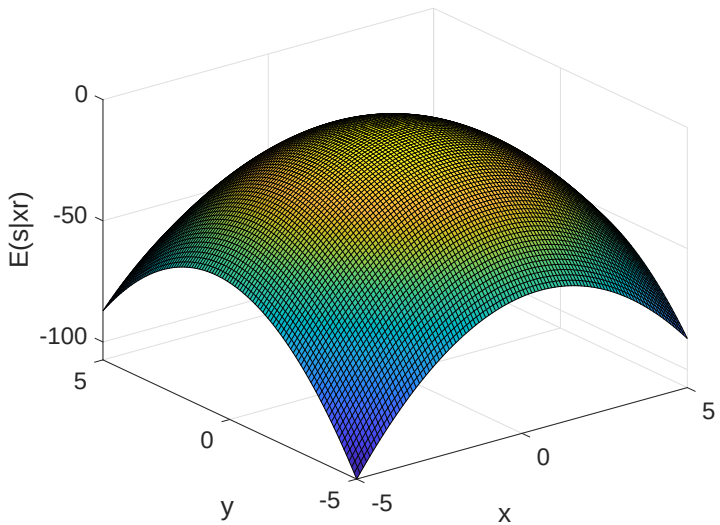
$$\left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B} \\ \hline \boldsymbol{C} & \boldsymbol{D} \end{array}\right] = \boldsymbol{Q}_1^{-1} \quad Z_1 = \boldsymbol{x}_r' \Sigma_t^{-1} \boldsymbol{x}_r \quad Z_2 = \boldsymbol{x}_r' \boldsymbol{A} \boldsymbol{x}_r$$

$$\mathbb{E}[s_{log} | \boldsymbol{x}_r] - const. = \operatorname{tr}(\boldsymbol{I}) - \operatorname{tr}(\boldsymbol{D}\Sigma_t) + Z_1 - Z_2$$

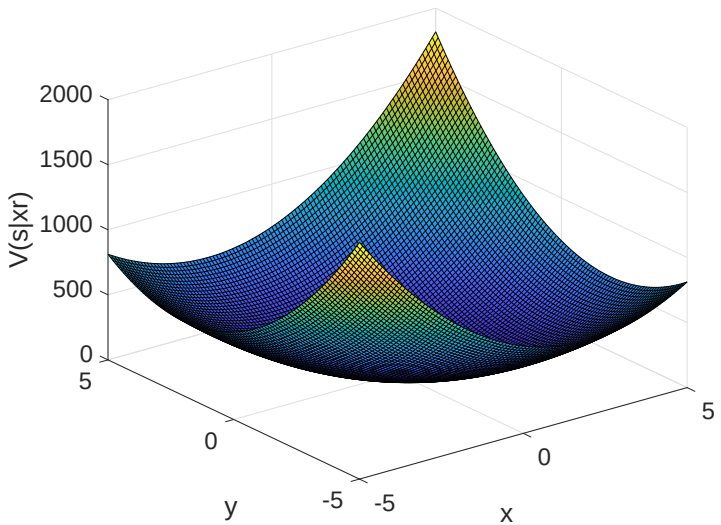
Fixed sample - variance

$$\begin{aligned}\mathbb{V}[s_{log}|\boldsymbol{x}_r] = \mathrm{tr}(\boldsymbol{W}_0\boldsymbol{\Sigma}_t)+ \\ + (\boldsymbol{x}_r'\boldsymbol{C})\boldsymbol{\Sigma}_t(\boldsymbol{x}_r'\boldsymbol{C})'+ \\ + (\boldsymbol{x}_r'\boldsymbol{B})\boldsymbol{\Sigma}_t(\boldsymbol{x}_r'\boldsymbol{B})'+ \\ + \mathrm{tr}(\boldsymbol{D}\boldsymbol{\Sigma}_t(\boldsymbol{D}+\boldsymbol{D}')\boldsymbol{\Sigma}_t)- \\ - 2\mathrm{tr}((\boldsymbol{D}+\boldsymbol{D}')\boldsymbol{\Sigma}_t)+ \\ + 2\mathrm{tr}(\boldsymbol{C}\boldsymbol{x}_r\boldsymbol{x}_r'\boldsymbol{B}\boldsymbol{\Sigma}_t)\end{aligned}$$

Fixed sample - example



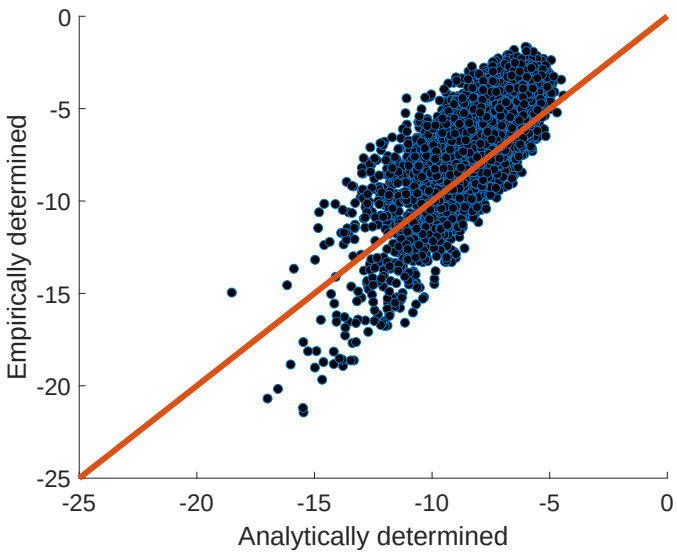
(a) Expectations for 2D samples and a given model.



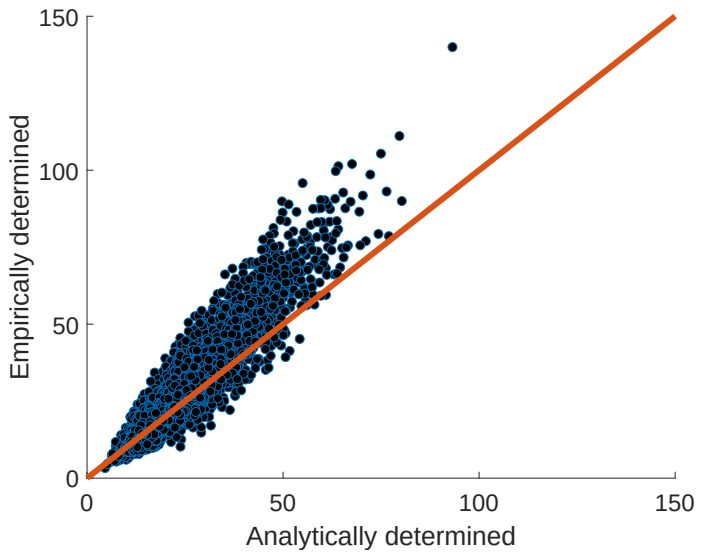
(b) Variances for 2D samples and a given model.

Difficulties in analytic normalization

Unfortunately, the questionable independence assumptions, non-normality of priors etc. all cause the ultimate failure of the analytic normalization approach, at least for the MOBIO database. However, the unequal exps. and vars. are a proven fact.



(a) Empirically vs. analytically determined expectations.



(b) Empirically vs. analytically determined variances.

Normalization techniques

Available literature proposes a few different normalization techniques (Petrovska-Delacrtaz, Chollet, and Dorizzi 2009), notably:

- *t*-norm (standardization for the test sample),
- *z*-norm (standardization for the reference sample),
- *tz*-norm,
- *zt*-norm,
- nonparametric CDF transformations.

Rationale behind normalization schemes is often described in a fuzzy and inconsistent way. Surprisingly, *zt* and *tz*-norm indeed **are** statistically correct, especially in an iterative process, when they are variants of **successive standardization**.

Successive standarization

It is possible to transform a rectangular array to have both row and column means 0 and std. deviations 1 (Olshen and Rajaratnam 2012).

Algorithm 1 Pseudocode for successive standardization of a $N \times M$ (at least 3×3) array $\boldsymbol{X}$.

1: **for** $i = 1$ to K **do**
2: **for** $n = 1$ to N **do**
3: $\boldsymbol{X}_{n,*} = \frac{\boldsymbol{X}_{n,*} - \bar{\boldsymbol{X}}_{n,*}}{s(\boldsymbol{X}_{n,*})}$
4: **end for**
5: **for** $m = 1$ to M **do**
6: $\boldsymbol{X}_{*,m} = \frac{\boldsymbol{X}_{*,m} - \bar{\boldsymbol{X}}_{*,m}}{s(\boldsymbol{X}_{*,m})}$
7: **end for**
8: **end for**

This trivial algorithm fails only for a set of Lebesgue measure 0.

Successive standarization - example

$$\boldsymbol{A}_0 = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 8 & 7 \end{bmatrix} \qquad \mu_r = \begin{bmatrix} 2.0 \\ 5.0 \\ 8.0 \end{bmatrix} \mu_c = \begin{bmatrix} 4.67 & 5.0 & 5.33 \end{bmatrix}$$

$$s_r = \begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \end{bmatrix} s_c = \begin{bmatrix} 4.04 & 3.0 & 2.08 \end{bmatrix}$$

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.95 & 0.09 & -1.04 \\ -0.88 & -0.21 & 1.09 \\ 0.81 & 0.30 & -1.12 \end{bmatrix} \qquad \mu_r = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} \mu_c = \begin{bmatrix} 0.29 & 0.06 & -0.36 \end{bmatrix}$$

$$s_r = \begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \end{bmatrix} s_c = \begin{bmatrix} 1.02 & 0.26 & 1.25 \end{bmatrix}$$

$$\boldsymbol{A}_8 = \begin{bmatrix} 1.15 & -0.63 & -0.53 \\ -0.63 & -0.53 & 1.15 \\ -0.53 & 1.15 & -0.63 \end{bmatrix} \qquad \mu_r = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} \mu_c = \begin{bmatrix} 0.0 & 0.0 & 0.0 \end{bmatrix}$$

$$s_r = \begin{bmatrix} 1.0 \\ 1.0 \\ 1.0 \end{bmatrix} s_c = \begin{bmatrix} 1.0 & 1.0 & 1.0 \end{bmatrix}$$

Experimental results - parametric

Additional results after the aggregation of normalized scores are provided. EERs for the DEV subset of MOBIO database (both genders at once):

Method	–	A-t	A-z	A-$(t+z)$
EER (before aggr.)	21.57	**21.45**	21.75	21.55
EER (after aggr.)	16.55	**16.30**	16.95	16.44

Method	–	t	z	tz	zt	$t+z$	S-tz	S-zt
EER (before aggr.)	21.57	**20.23**	21.60	20.43	20.62	21.18	20.40	20.45
EER (after aggr.)	16.55	**13.95**	17.11	14.90	14.90	15.91	14.90	14.70

Experimental results - nonparametric

All distributions were transformed into standard normal.

Method	–	*nt*	*nz*	*ntz*	*nzt*	*nt* + *nz*	S-*ntz*	S-*nzt*
EER (before aggr.)	21.57	**20.48**	21.69	20.53	20.89	20.33	20.91	20.86
EER (after aggr.)	16.55	**13.66**	17.01	13.76	14.73	14.25	14.44	14.59

Experimental results - selected ROCs after aggregation

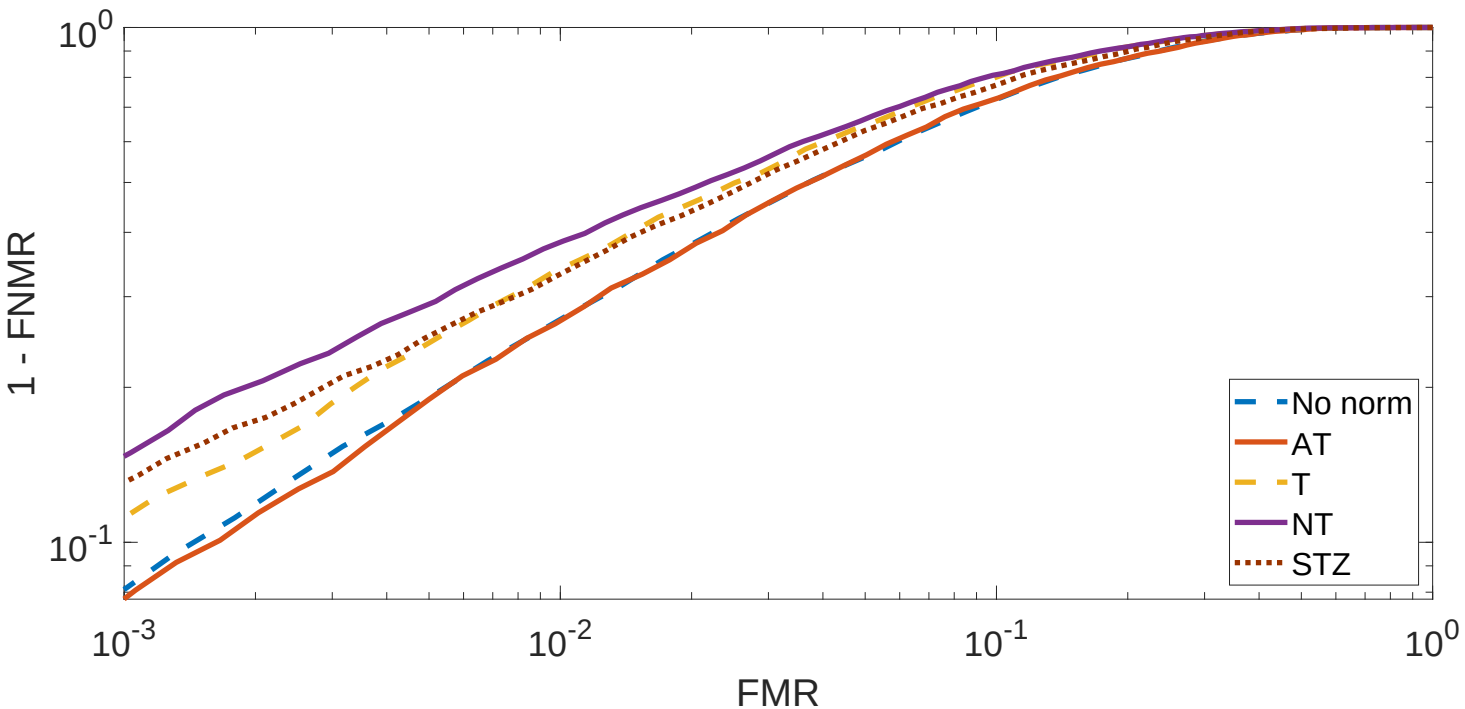


Figure: ROC curves for selected normalization schemes.

Summary

Unfortunately, even for simple PLDA models it is difficult to create a general normalization scheme using purely closed-form, analytic calculations. Nevertheless, it's now been proven that the distributions of scores for various identities and samples are **indeed different.**

The best results were achieved using nonparametric *nt*-norm scheme but it's strongly encouraged that every case should be investigated individually. Successive methods can be very efficient in the low-FMR areas.

Future work

Future work will be concentrated on:

- getting more (**a lot more**) empirical data,
- investigating the linear model further:
  - deriving the formulas for 2 or more factor matrices,
  - assuming other (heavy-tailed) priors: $\boldsymbol{h}_r | u_h \sim \mathcal{N}(\mathbf{0}, u_h^{-1}\boldsymbol{I})$ where $u_h \sim \mathcal{G}(n/2, n/2)$,
- trying other nonparametric methods (like in: Štruc, Gros, and Pavešić 2012).

Bibliography


Olshen, Richard A. and Bala Rajaratnam (2012). "Successive Standardization of Rectangular Arrays". In: *Algorithms* 5.1, pp. 98–112. DOI: 10.3390/a5010098.


Petersen, K. B. and M. S. Pedersen (2012). *The Matrix Cookbook*. Version 20121115. URL: http://www2.imm.dtu.dk/pubdb/p.php?3274.


Petrovska-Delacrtaz, Dijana, Grard Chollet, and Bernadette Dorizzi (2009). *Guide to Biometric Reference Systems and Performance Evaluation*. 1st ed. Springer Publishing Company, Incorporated. ISBN: 1848002912, 9781848002913.


Prince, S. J. D. and J. H. Elder (2007). "Probabilistic Linear Discriminant Analysis for Inferences About Identity". In: *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8. DOI: 10.1109/ICCV.2007.4409052.


Štruc, V., J. Ž. Gros, and N. Pavešić (2012). "Non-parametric score normalization for biometric verification systems". In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 2395–2399.


Yager, Neil and Ted Dunstone (2010). "The Biometric Menagerie". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 32.2, pp. 220–230. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2008.291.