

# Estimating conditional density of missing values using deep Gaussian mixture model

Marcin Przewięźlikowski

Marek Śmieja

Łukasz Struski

We propose **Deep Mixture of Factor Analyzers** (DMFA) for estimating the conditional probability density function of missing values.

## DMFA:

- combines the features of deep learning models and GMMs.
- follows a discriminative approach and directly maximizes the likelihood function **on missing values**.
- estimates a covariance matrix in the form of Factor Analysers:  $\Sigma = AA^T + D$ , which works well for high dimensional data.

Code available at:

[https://github.com/mprzewie/dmfa\\_inpainting](https://github.com/mprzewie/dmfa_inpainting)

Faculty of Mathematics and Computer Science, Jagiellonian University, Łojasiewicza 6,  
30-428 Kraków, Poland. Correspondence to: Marek Śmieja <marek.smieja@uj.edu.pl>



JAGIELLONIAN UNIVERSITY  
IN KRAKÓW

ICML | 2020

Thirty-seventh International Conference  
on Machine Learning

ARTEMIS 2020

group of machine  
**gmum**  
learning research

# Deep conditional Gaussian density for missing data

- The proposed DMFA creates a Gaussian density, which minimizes the negative log likelihood on missing values.
- During training, given a data point, we first generate a random binary mask  $(x, M)$  to simulate missing attributes. The pair induces a missing data point  $(x_o, x_m)$ .
- We define conditional Gaussian density as

$$p(x_m|x_o) = N(\mu_m, \Sigma_{mm})$$

where  $\mu$  and  $\Sigma$  are the output of DMFA model given  $(x_o, x_m)$  and  $\mathbf{M}$ , and  $(\mu_m, \Sigma_{mm})$  denotes the restriction of  $\mu$  and  $\Sigma$  to coordinates of missing data.

- DMFA does not estimate the density of the **entire** data, but is trained to minimize directly:  
 $l(x_o, x_m) = -\log p(x_m|x_o) = -\log N(\mu_m, \Sigma_{mm})(x_m)$   
and should thus provide a better estimate of  $x_m$  than conditional density obtained by a typical GMM.

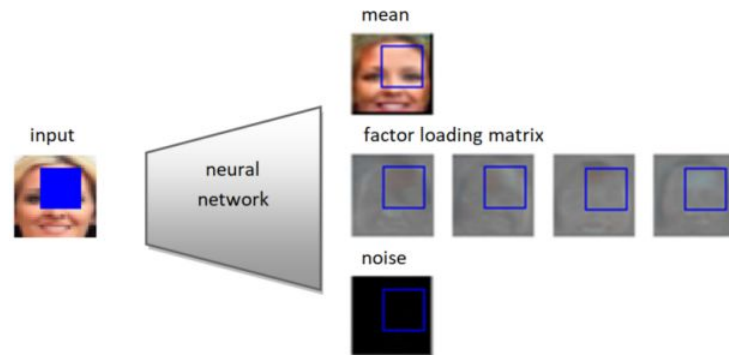


Figure 1. The idea of the proposed DMFA. Given a missing data point, our model returns the parameters of conditional Gaussian density: mean, factor loading matrix (we use 4 latent factors) and noise matrix for the model of Factor Analyzers, which describes a distribution of missing data (the area inside the blue square).

# Experiments

We compared imputations produced by DMFA and conventional MFA on MNIST and CelebA datasets.

Table 1. Negative log-likelihood (NLL) and mean-square error (MSE) of the most probable imputation obtained by DMFA and MFA (lower is better).

Dataset	Measure	MFA	DMFA
MNIST	NLL	58.10	-244.81
	MSE	18.59	12.96
CelebA	NLL	-882.54	-1222.85
	MSE	9.82	7.73

Table 2. Negative log-likelihood (NLL) and mean-square error (MSE) of the most probable imputation obtained by DMFA and MFA when the sizes of missing regions were different in training and test set (test set contains the missing regions of the size: 10x15 for MNIST and 12x20 for CelebA).

Dataset	Measure	MFA	DMFA
MNIST	NLL	5.85	-147.29
	MSE	12.09	9.04
CelebA	NLL	-815.60	-885.85
	MSE	11.73	10.03

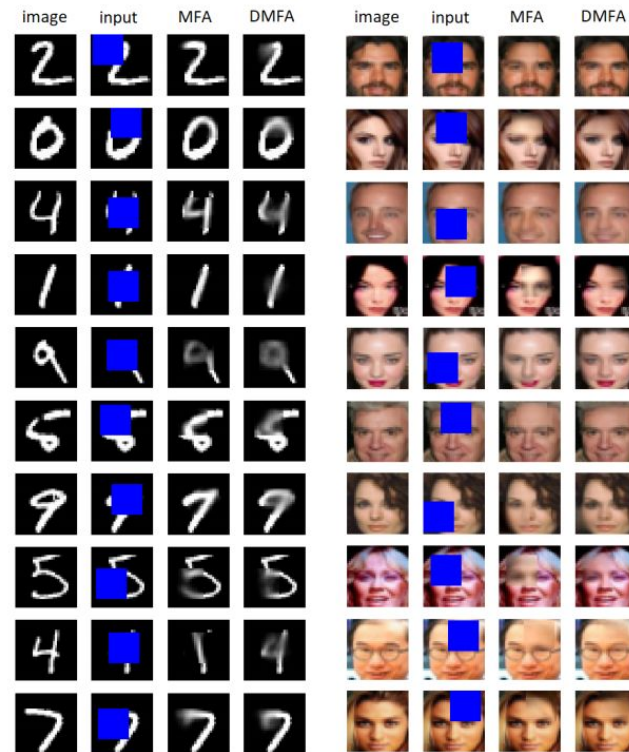


Figure 2. Sample imputation results produced by DMFA and MFA.

