Experimental Report
Mainly Based on DNF Models

2020.10.27

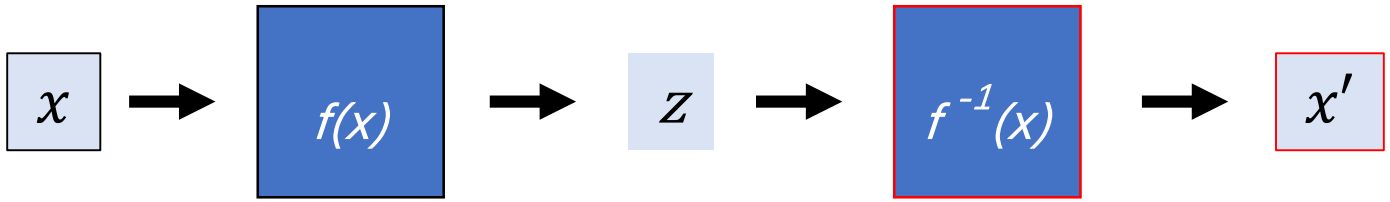
Hanjiao

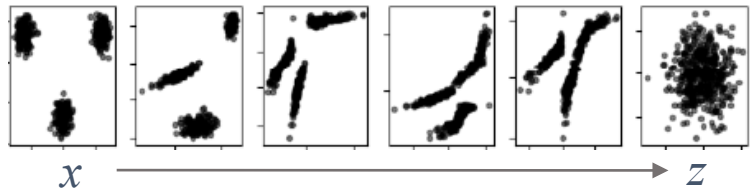
Introduction

- NF , DNF and DNF-MG
- DNF-MG experiments
- Sub-DNF experiments
- DNF experiment-verification

Flow-based Generative model

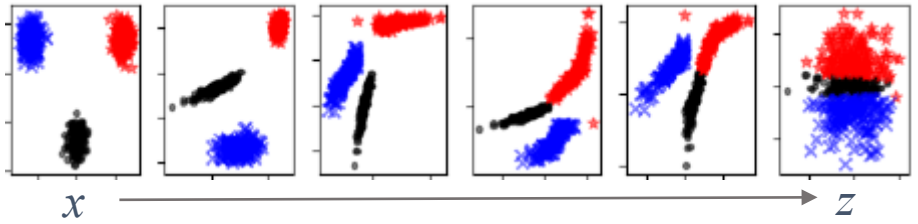
- flow





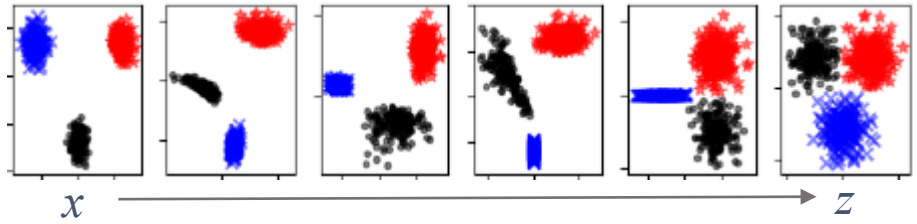
Normalization Flow (NF) *vs.* Discriminative Normalization Flow (DNF)

- NF



$$L(\boldsymbol{\theta}) = \sum_i \ln p(\mathbf{x}_i) = \sum_i \ln p(\mathbf{z}_i) + \sum_i \sum_{t=1}^{T+1} \ln \left| \det \frac{\partial f_{t-1}^{-1}(\mathbf{z}_{it})}{\partial \mathbf{z}_{it}} \right|$$

- DNF



$$p_y(\mathbf{z}) = N(\mathbf{z}; \boldsymbol{\mu}_y, \boldsymbol{\Sigma})$$

$$L(\boldsymbol{\Theta}) = \sum_i \ln(p_{y(\mathbf{x_i})}(\mathbf{z_i})) + \ln \left| \det \frac{\partial f^{-1}(\mathbf{x_i})}{\partial \mathbf{x_i}} \right|$$

DNF with Maximum Gaussianality(MG) Training

- Maximum Gaussianality (MG) training approach maximizes the Gaussianality of the latent codes directly.

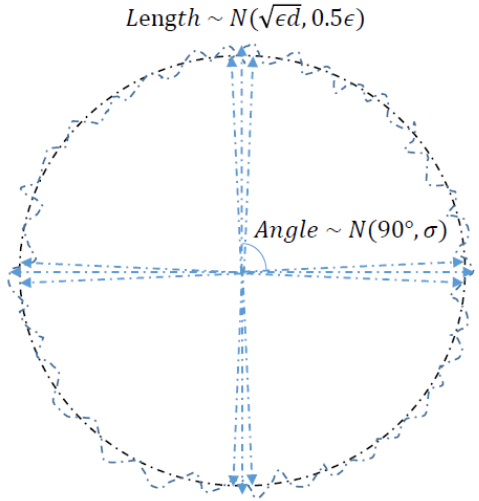


Fig. 6: Length and angle metrics of a high-dimensional Gaussian $N(0, \epsilon)$. The length of the samples can be approximated by a Gaussian $N(\sqrt{\epsilon d}, 0.5\epsilon)$, and the angle of two samples can be approximated by a Gaussian $N(90°, \sigma)$.

$$\mathcal{R}_\ell = -\sum_i ||\ell(\mathbf{x}_i) - \sqrt{\epsilon d}||^2$$

$$\mathcal{R}_\phi = -\sum_i \sum_j \frac{||\phi(\mathbf{x}_i, \mathbf{x}_j)||^2}{2\xi}$$

$$\mathcal{L}(\boldsymbol{\theta}) = \mathcal{R}_\ell + \mathcal{R}_\phi = -\sum_i ||\ell(\mathbf{z}_i) - \sqrt{d}||^2 - \sum_i \sum_j \frac{||\phi(\mathbf{z}_i, \mathbf{z}_j)||^2}{2\xi}$$

DNF with Maximum Gaussianality(MG) Training Experiments

- Datasets
  - VoxCeleb ： contains 2000+ hours of speech signals from 7000+ speakers.
  - SITW ： consists of 299 speakers.

- Model settings
  - x-vector
  - Normalization models ： based on the DNF architecture
  - Training criterion： MG training
  - Scoring models ： Cosine scoring and PLDA scoring

DNF with Maximum Gaussianality(MG) Training Experiments

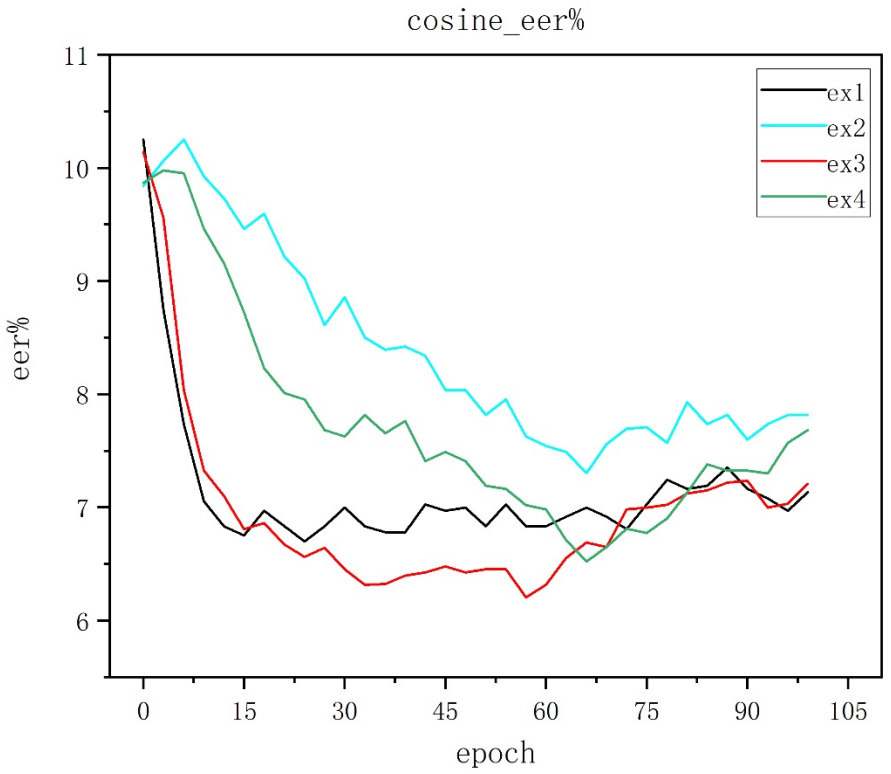
- Note
  - ex1, ex2 (n-filter = 50) : use 1200+ speakers with more than 50 utterances.
    ex3, ex4 (n-filter = 10) : use 4000+ speakers with more than 10 utterances.
  - total angle loss
    within-class angle loss

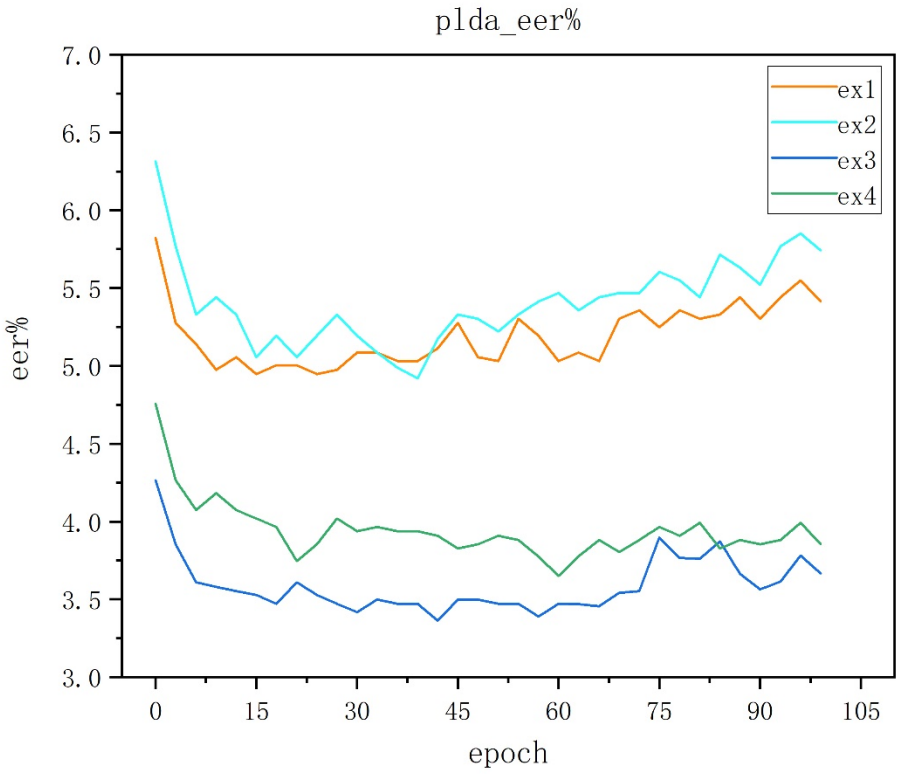
EX	n-filter	Model	Parameters				EER%	
			L2-sb	cos-sb	L2-sw	cos-sw	Cosine	PLDA
ex1	50	total angle loss	10	500	10	10	6.658	4.948
ex2	50	within-class angle loss	10	500	10	10	7.302	4.920
ex3	10	total angle loss	10	500	10	10	**6.203**	**3.362**
ex4	10	within-class angle loss	10	500	10	10	6.522	3.650

- The number of utterances per speaker can affect performance.

DNF with Maximum Gaussianality(MG) Training Experiments

- EER% results on SITW with x-vector frontend.





Subspace-DNF Experiments-1

- Datasets
  - VoxCeleb ： only use 4000+ speakers with more than 10 utterances.
  - SITW

- Model settings
  - x-vector
  - Normalization models ： LDA model
  - Dimensions ： original x-vector with 512 dimensions and x-vector with 512 plus 100-dimensional Gaussian noise.
  - Scoring models ： Cosine scoring and PLDA scoring

- EER(%) results on SITW of x-vector frontend and x-vector with 100-dimensional Gaussian noise as well as their LDA results respectively.

Model		**Sitw-eval**	
		EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[**512**]	17.2	5.303
Linear	LDA[512]	8.611	5.303
	LDA[400]	7.381	4.647
	LDA[200]	5.823	**3.964**
	LDA[150]	5.249	4.073
	LDA[100]	4.811	4.101
	LDA[60]	**4.429**	4.483
	LDA[50]	4.921	4.948

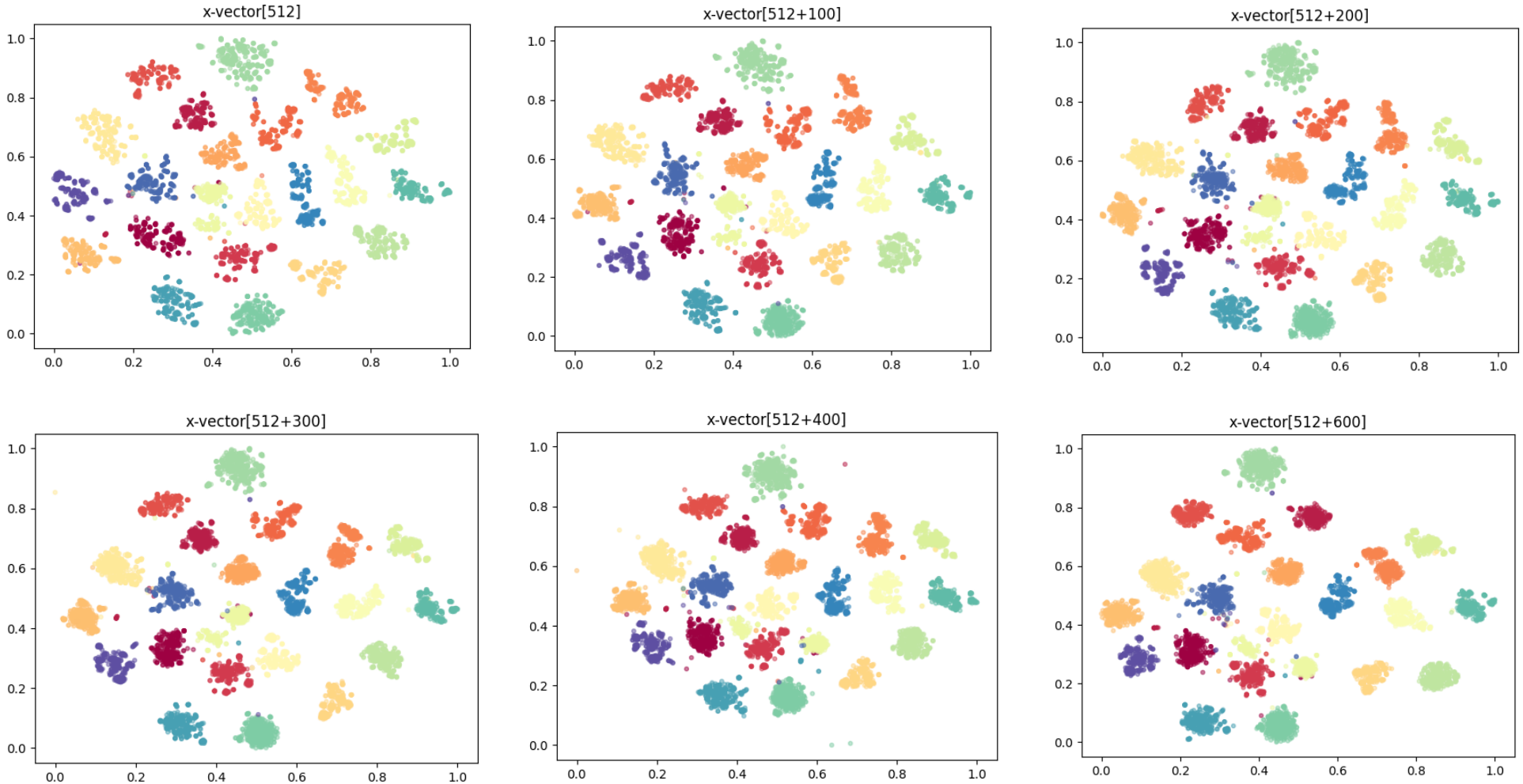
Model		**Sitw-eval**	
		EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[**512+100**]	17.52	4.975
Linear	LDA[512]	8.502	5.221
	LDA[400]	7.354	4.62
	LDA[200]	5.768	**3.937**
	LDA[150]	5.249	4.128
	LDA[100]	4.839	4.046
	LDA[60]	**4.429**	4.485
	LDA[50]	4.893	4.921

- EER(%) results on SITW of x-vector frontend and x-vector with different dimensional Gaussian noise.

Model		**Sitw-eval**	
		EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[**512**]	17.2	5.303
	x-vector[**512+50**]	17.44	5.085
	x-vector[**512+100**]	17.52	4.975
	x-vector[**512+150**]	17.61	4.866
	x-vector[**512+200**]	17.99	4.675
	x-vector[**512+300**]	18.04	**4.647**
	x-vector[**512+350**]	18.15	4.839
	x-vector[**512+400**]	18.02	4.675
	x-vector[**512+450**]	18.34	4.757
	x-vector[**512+600**]	18.75	4.921

- As the dimensionality of Gaussian noise increases, PLDA scoring models show better performance.

- t-SNE of original x-vector and x-vector with different dimensional Gaussian noise.



Subspace-DNF Experiments-2

- EER% results on SITW using sub-space DNF models.
- Table 2 is for comparison.

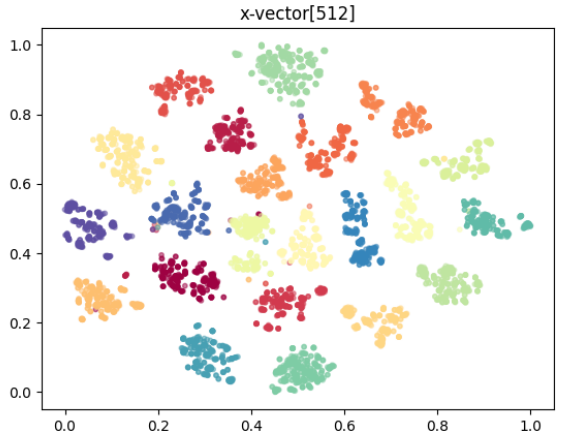
Model		**Sitw-eval**	
		EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
Nonlinear	DNF[**512**]	8.66	3.69
	DNF[**512+100**]	10.25	3.8
	DNF-S[512]	9.896	3.909
	DNF-S[400]	9.732	3.882
	DNF-S[150]	9.049	4.893
	DNF-S[100]	9.322	5.768
	DNF-S[50]	11.15	8.283

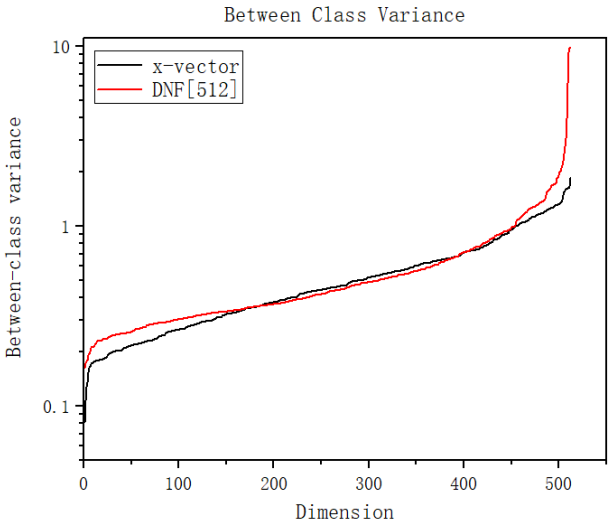
Table 1

Model		**Sitw-eval**	
		EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[**512+100**]	17.52	4.975
Linear	LDA[512]	8.502	5.221
	LDA[400]	7.354	4.62
	LDA[200]	5.768	**3.937**
	LDA[150]	5.249	4.128
	LDA[100]	4.839	4.046
	LDA[60]	**4.429**	4.485
	LDA[50]	4.893	4.921

Table 2

- t-SNE of the original x-vectors with 512 dimensions and the latent codes produced by DNF model.



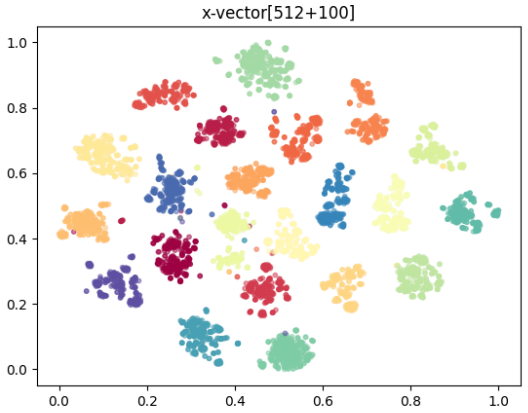


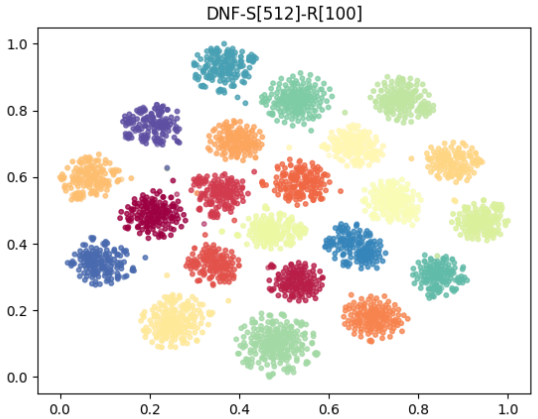
Between-class variation

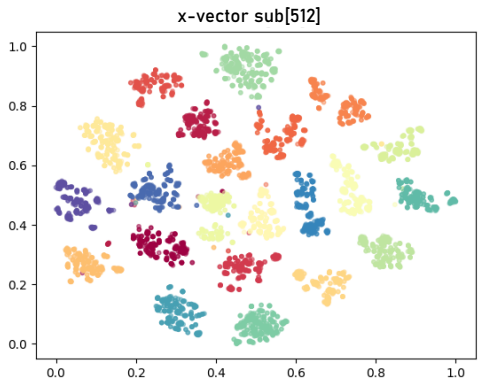
Computed on the original x-vectors and the latent codes produced by DNF.

- t-SNE of x-vector with 100-dimensional Gaussian noise(left) and the latent codes produced by sub-space DNF model(right).

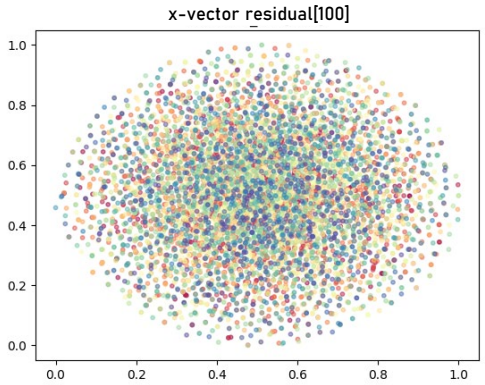
Full dimensions



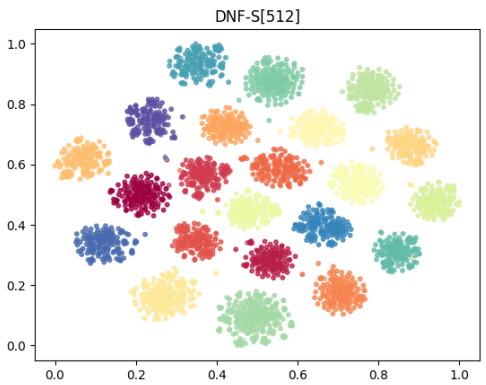


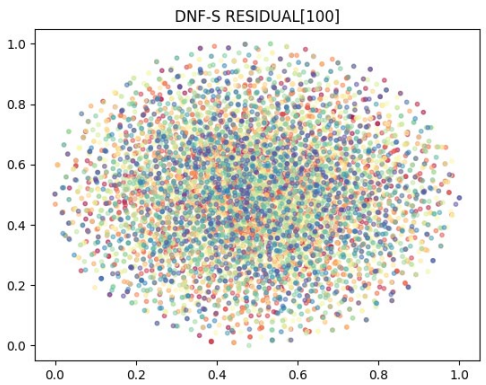


sub-space dimensions



residual space dimensions





DNF experiment-verification

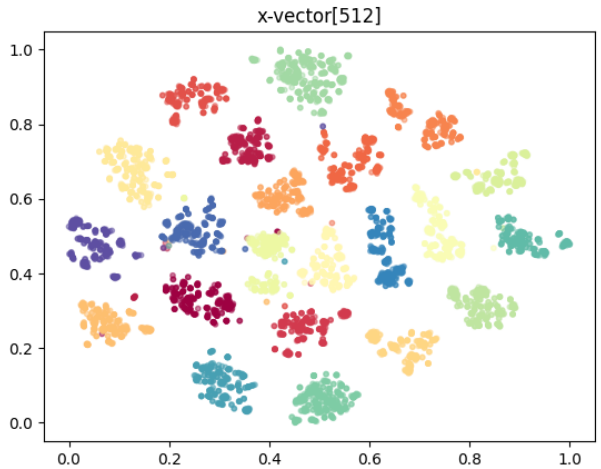
- DNF experiment-verification

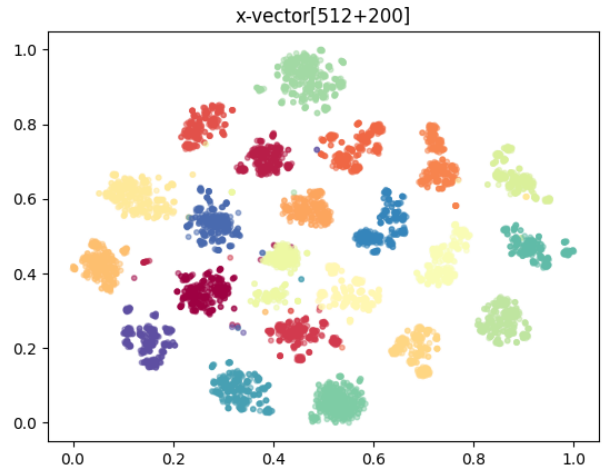
xvector[512]	
SITW	
Cosine	**Plda**
17.61	5.303

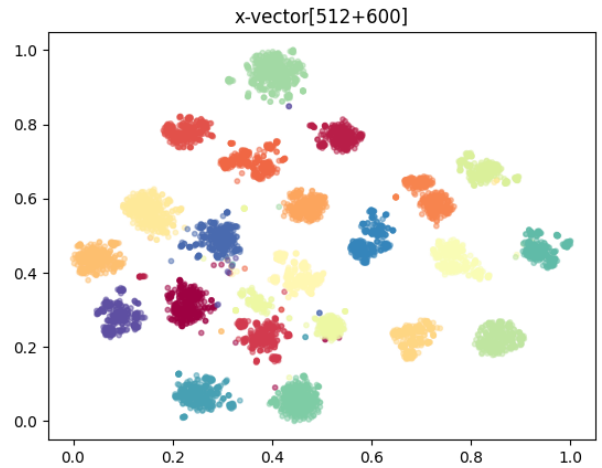
xvector[512+100]	
SITW	
Cosine	**Plda**
17.61	4.893
17.61	4.839
17.47	5.03
17.47	5.03
17.5	4.921
17.58	4.866
17.44	4.921
17.52	5.003
17.44	4.921
17.55	4.893

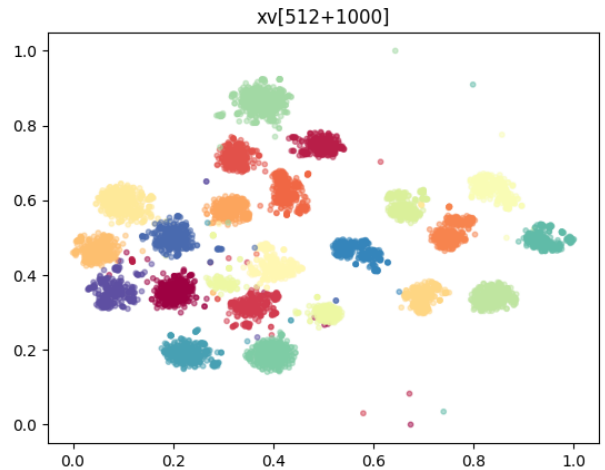
- Ten groups of similar scoring results show that PLDA scoring performance has indeed improved.

- DNF experiment-verification









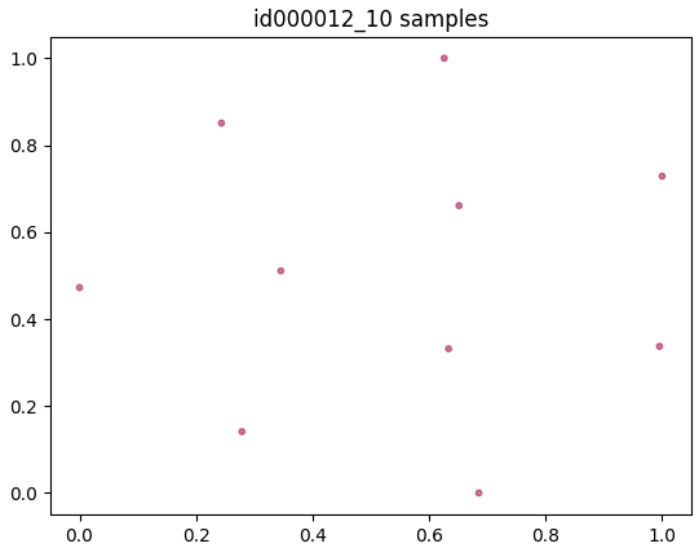
- DNF experiment-verification

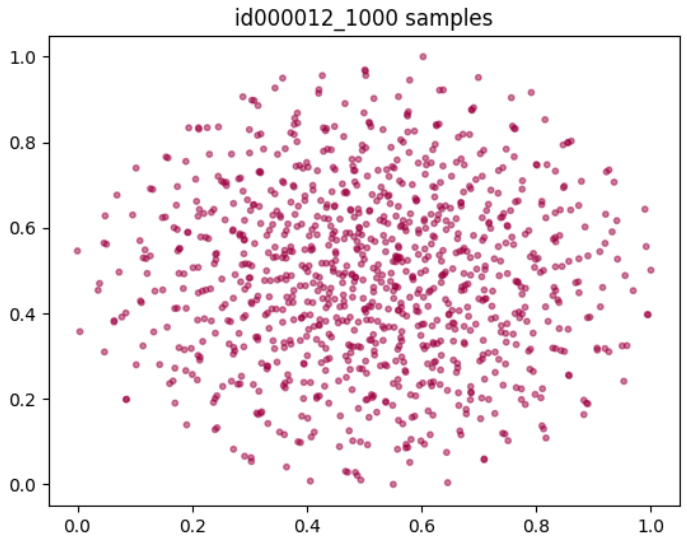
Model		Sitw-eval EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[512+100]	17.52	4.975
Linear	LDA[512]	8.502	5.221
	LDA[400]	7.354	4.62
	LDA[200]	5.768	**3.937**
	LDA[150]	5.249	4.128
	LDA[100]	4.839	4.046
	LDA[60]	**4.429**	4.485
	LDA[50]	4.893	4.921

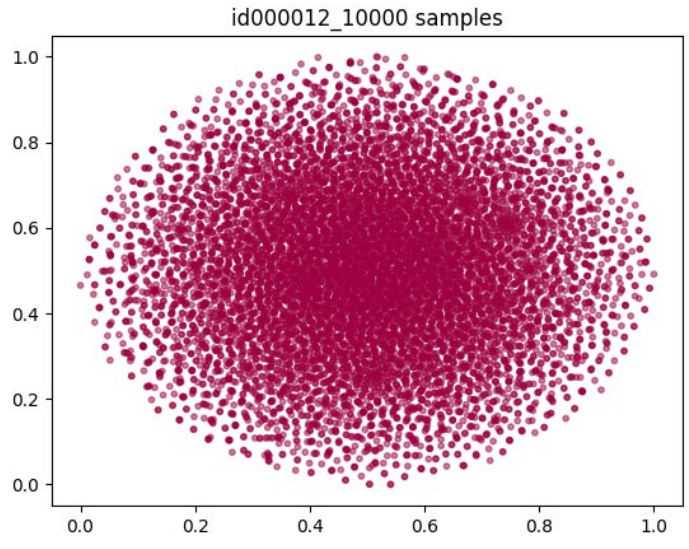
Model		Sitw-eval EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
Nonlinear	DNF[512]	8.66	3.69
	DNF[512+100]	10.25	3.8
	DNF-S[512]	9.896	3.909
	DNF-S[400]	9.732	3.882
	DNF-S[150]	9.049	4.893
	DNF-S[100]	9.322	5.768
	DNF-S[50]	11.15	8.283

Model		Sitw-eval EER(%)	
Dim_reduction	System[Dim]	Cosine	PLDA
N/A	x-vector[512]	17.2	5.303
	x-vector[512+50]	17.44	5.085
	x-vector[512+100]	17.52	4.975
	x-vector[512+150]	17.61	4.866
	x-vector[512+200]	17.99	4.675
	x-vector[512+300]	18.04	**4.647**
	x-vector[512+350]	18.15	4.839
	x-vector[512+400]	18.02	4.675
	x-vector[512+450]	18.34	4.757
	x-vector[512+600]	18.75	4.921
	x-vector[512+1000]	18.89	5.085

- DNF experiment-verification







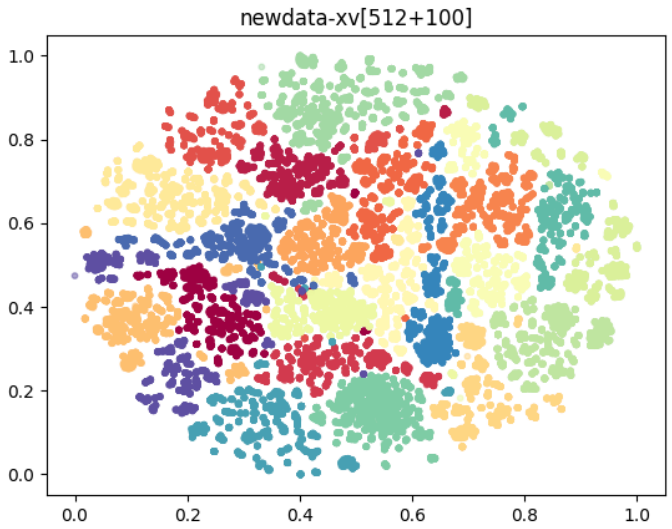
- DNF experiment-verification

  - original dataset : voxceleb , 19w+ samples , 4000+ speakers , 512 dims.
  - dim extended : voxceleb , 19w+ samples , 4000+ speakers , 612 dims.
  - data extended : voxceleb, 190w+ samples, 4000+ speakers , 612 dims.

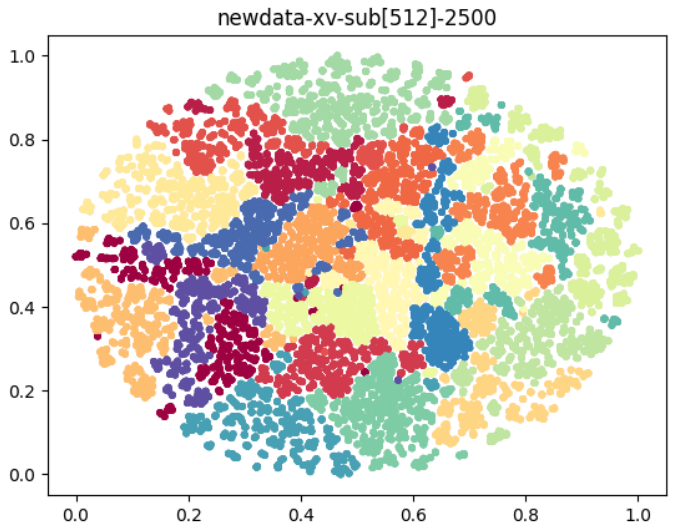
ex	original dataset xvector-dim[512]		dim extended xvector-dim[512+100]		data extended xvector-dim[512+100]	
	SITW		SITW		SITW	
	Cosine	Plda	Cosine	Plda	Cosine	Plda
1	17.2	5.303	17.52	4.975	17.55	4.921

- DNF experiment-verification

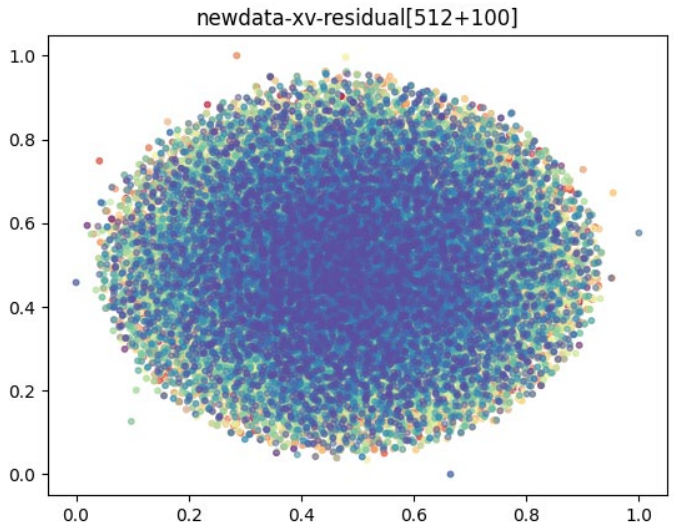
  - new data : voxceleb, 190w+ samples, 4000+ speakers , 612 dims.
  - tsne : select speakers with utts larger than 2500.



full dimension



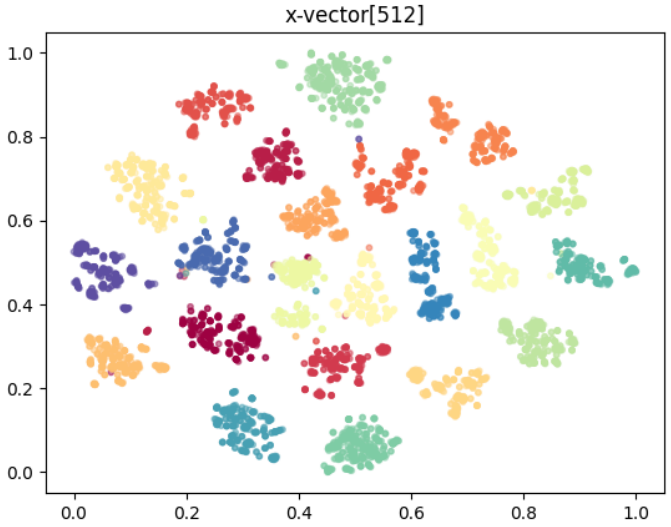
subspace



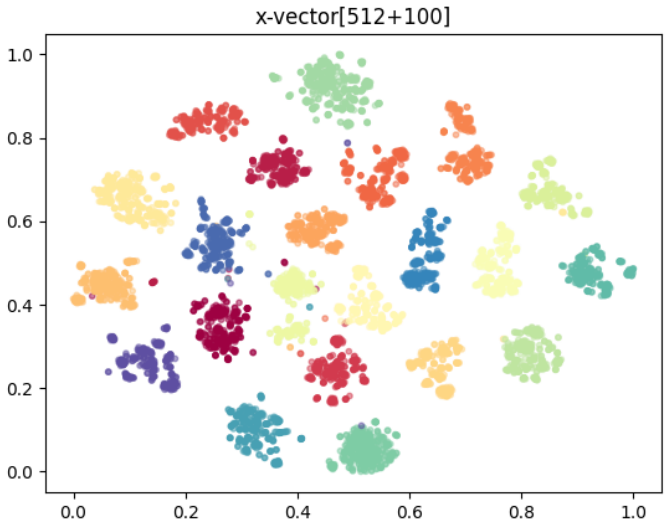
residual space

- DNF experiment-verification

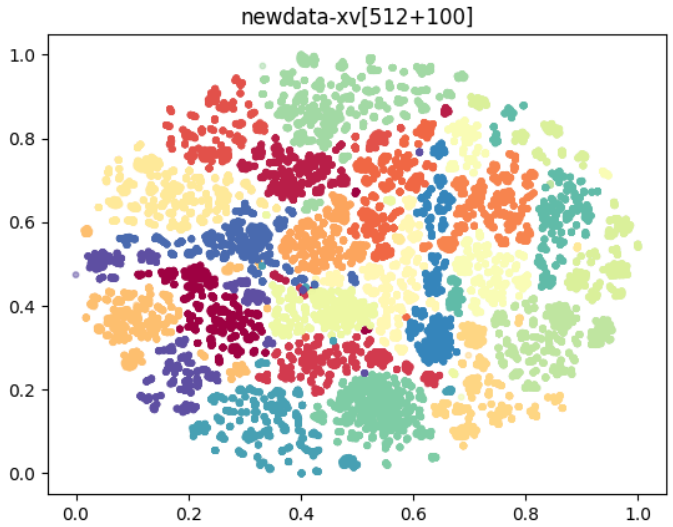
  - original dataset : voxceleb, 19w+ samples, 4000+ speakers .
  - new dataset : voxceleb, 190w+ samples, 4000+ speakers , 612 dims.



original dataset
512dims

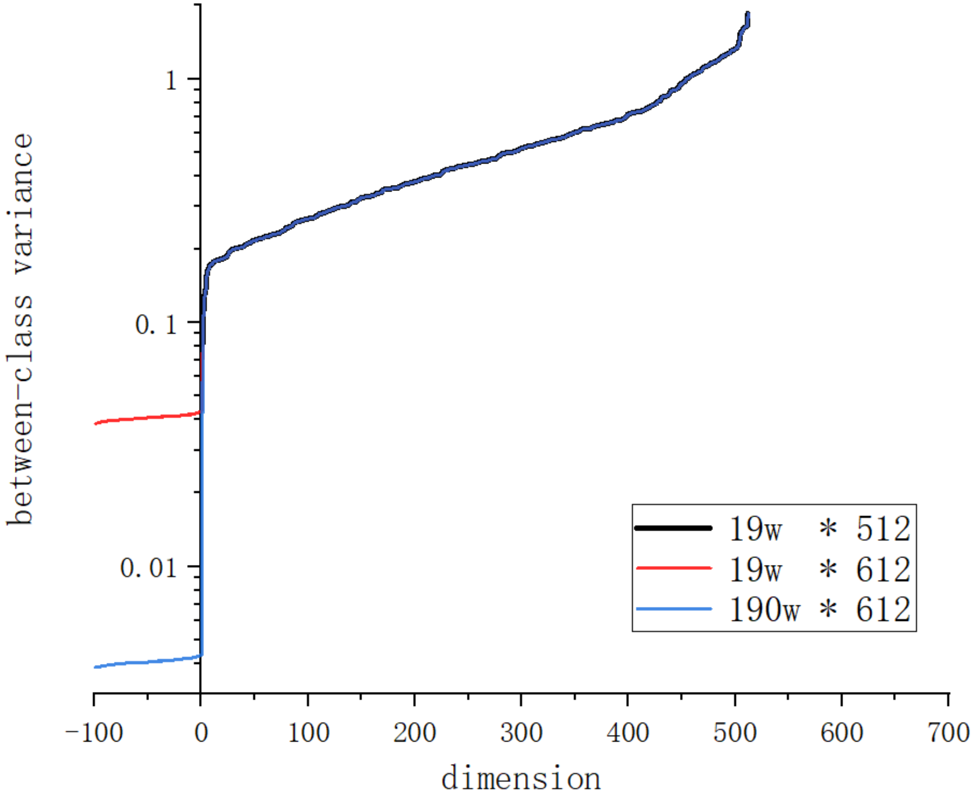


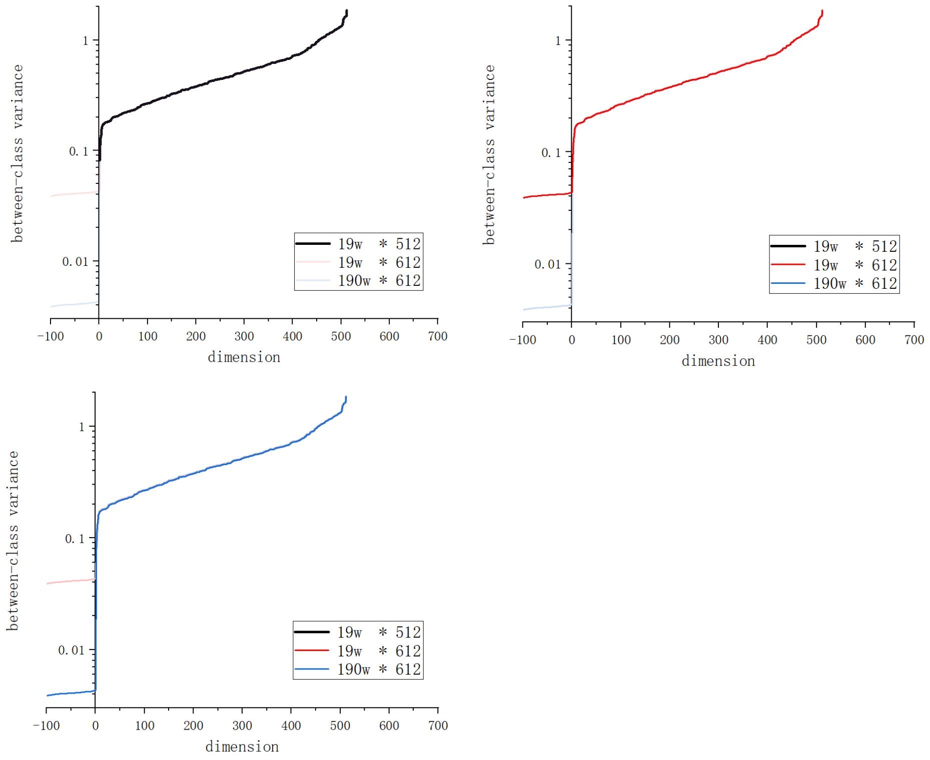
original dataset
612dims



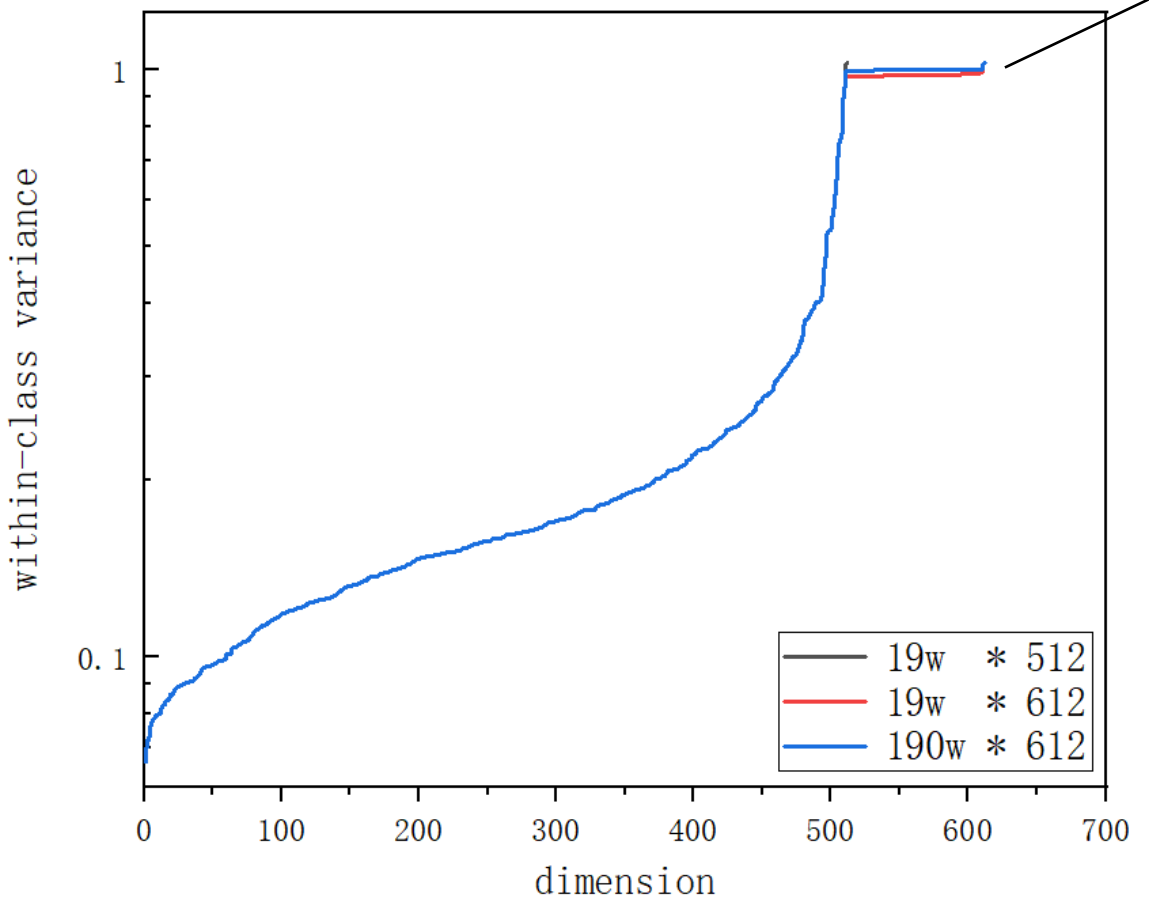
new dataset
612 dims

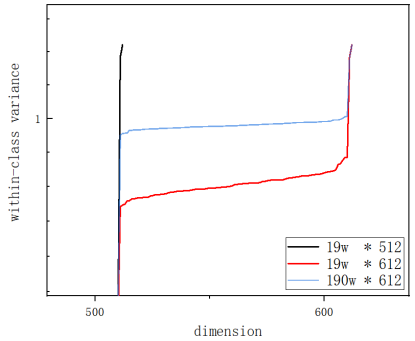
- DNF experiment-verification
  - Comparison of between-class variance

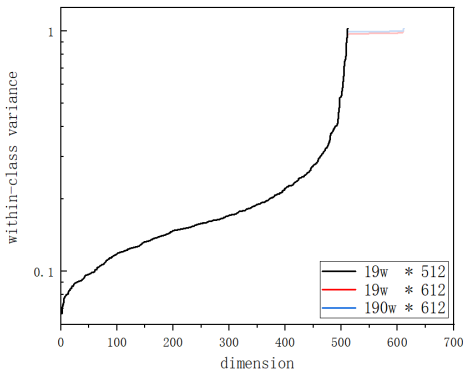


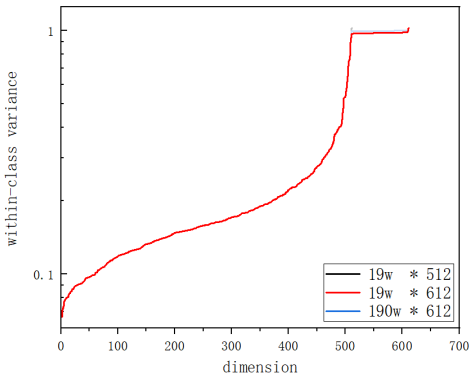


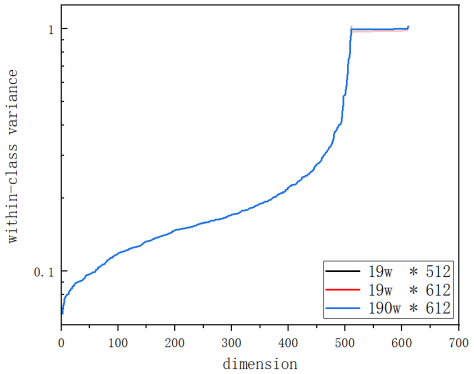
- DNF experiment-verification
  - Comparison of within-class variance











- DNF experiment-verification

	length norm / normalize length	512 dims	512dims +100 gaussian noise					512dims +400 gaussian noise	512dims +1000 gaussian noise
			ex 1	ex 2	ex 3	ex 4	ex 5		
Plda	T / T	5.44	5.057	5.112	5.139	5.139	5.139	4.866	5.167
	T / F	6.37	6.151	6.26	6.37	6.151	6.178	6.151	6.37
	F / T	**5.303**	**4.893**	**4.921**	**5.03**	**4.839**	**5.03**	**4.729**	**5.085**
	F / F	6.506	6.506	6.534	6.506	6.534	6.561	6.588	6.616
Cosine		17.2	17.61	17.47	17.5	17.61	17.47	18.43	19.3

- DNF experiment-verification
  - The EER(%) results for different types of noise added on each dataset(vox_4k , enroll , test).
  - Noise : sampled from normal distributions with different variances and different constants.
  - Cosine : --use global mean / --not use global mean
  - The scoring results are same with noise as a constant.

	length norm / normalize length	**dim[512]**	**dim[512+100]**					
			noise *N*(0, 1)	**noise *N*(0, 0.1)**	**noise *N*(0, 10)**	**noise 0**	**noise 1**	**noise 10**
Plda	T / T	5.44	5.057	5.194	**4.839**	5.44	5.44	5.44
	T / F	6.37	6.151	6.397	7.053	6.37	6.37	6.37
	F / T	**5.303**	**4.893**	**4.948**	4.948	**5.303**	**5.303**	**5.303**
	F / F	6.506	6.506	6.561	6.534	6.506	6.506	6.506
Cosine		17.2/16.79	17.61/16.89	17.2/16.81	46.2/46.45	17.2/16.79	17.2/16.79	17.2/16.79

- DNF experiment-verification

  - The EER(%) results for different types of noise added on training dataset(vox_4k) while the last 100 dimensions of enroll set and test set are all 0.

  - Noise : sampled from normal distributions with different variances and different constants.

  - Cosine : --use global mean / --not use global mean

	length norm / normalize length	**dim[512]**	**dim[512+100]**				
			trainset noise *N(0, 1)*	**trainset noise *N(0, 0.1)***	**trainset noise *N(0, 10)***	**trainset noise 1**	**trainset noise 10**
Plda	T / T	5.44	5.413	5.44	**5.194**	99.97	99.97
	T / F	6.37	6.37	6.37	6.534	44.23	44.61
	F / T	**5.303**	**5.276**	**5.303**	5.303	99.97	99.97
	F / F	6.506	6.506	6.506	6.506	99.97	99.97
Cosine		17.2/16.79	17.06/16.79	17.2/16.79	16.68/16.79	16.95/16.79	16.76/16.79

Updating…