OpenMax with Clustering for Open-Set Classification

Dennys Huber

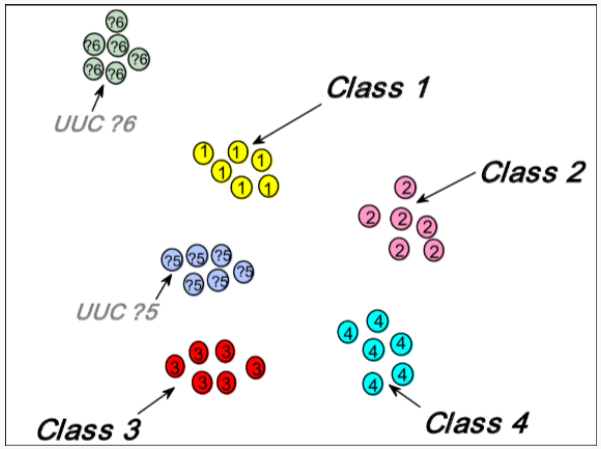
February 5, 2024

University of Zurich

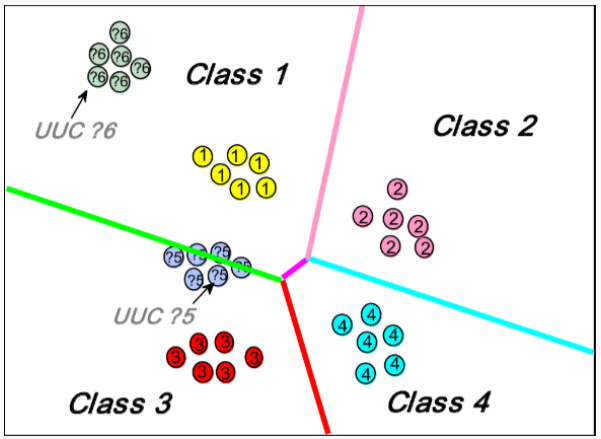
Table of contents

Open-Set

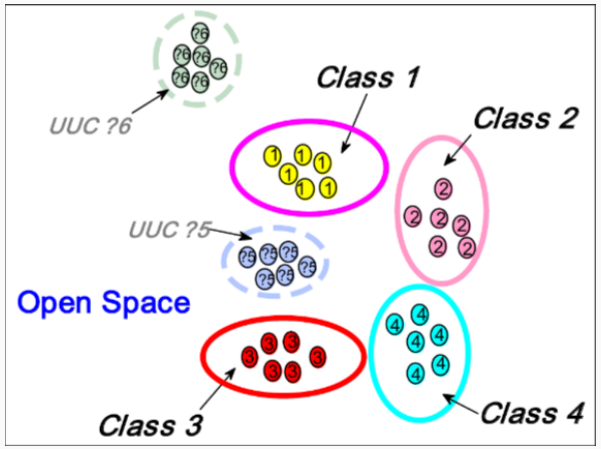
Closed-Set Problem



(a) Distribution of the original dataset



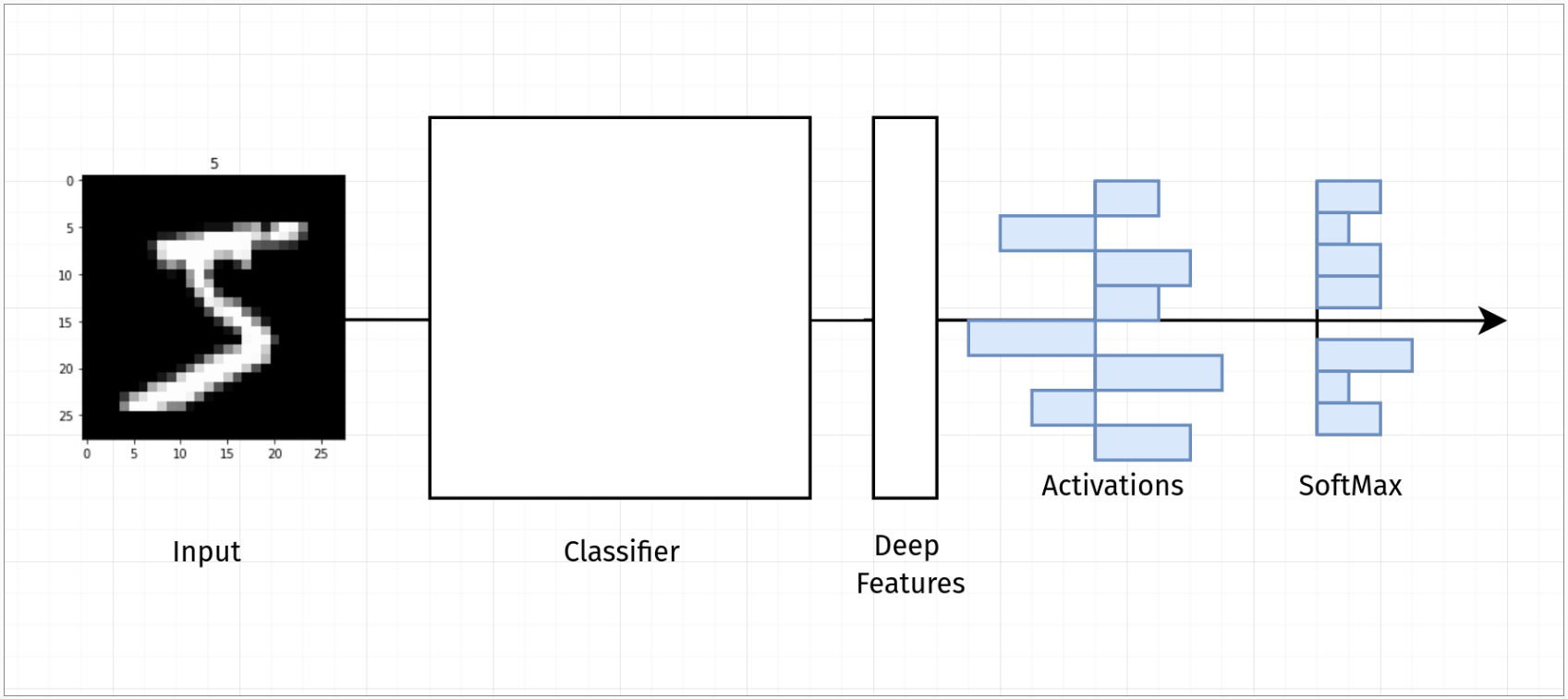
(b) Closed-set classification problem



(c) Open-set classification problem

OpenMax

SoftMax



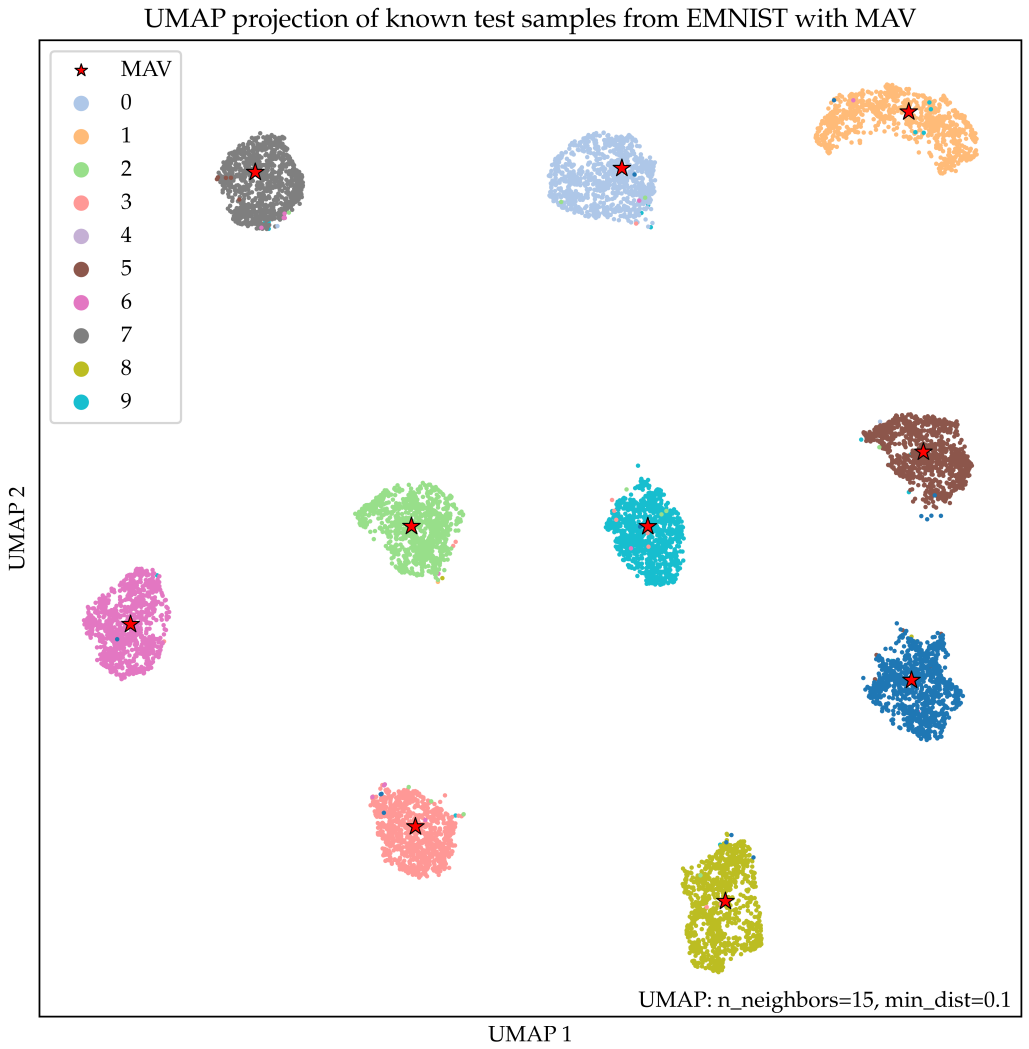
Introducing OpenMax

- Extension of SoftMax during testing
- Distance based
- Deep features
- Extreme Value Theory
- Heuristic

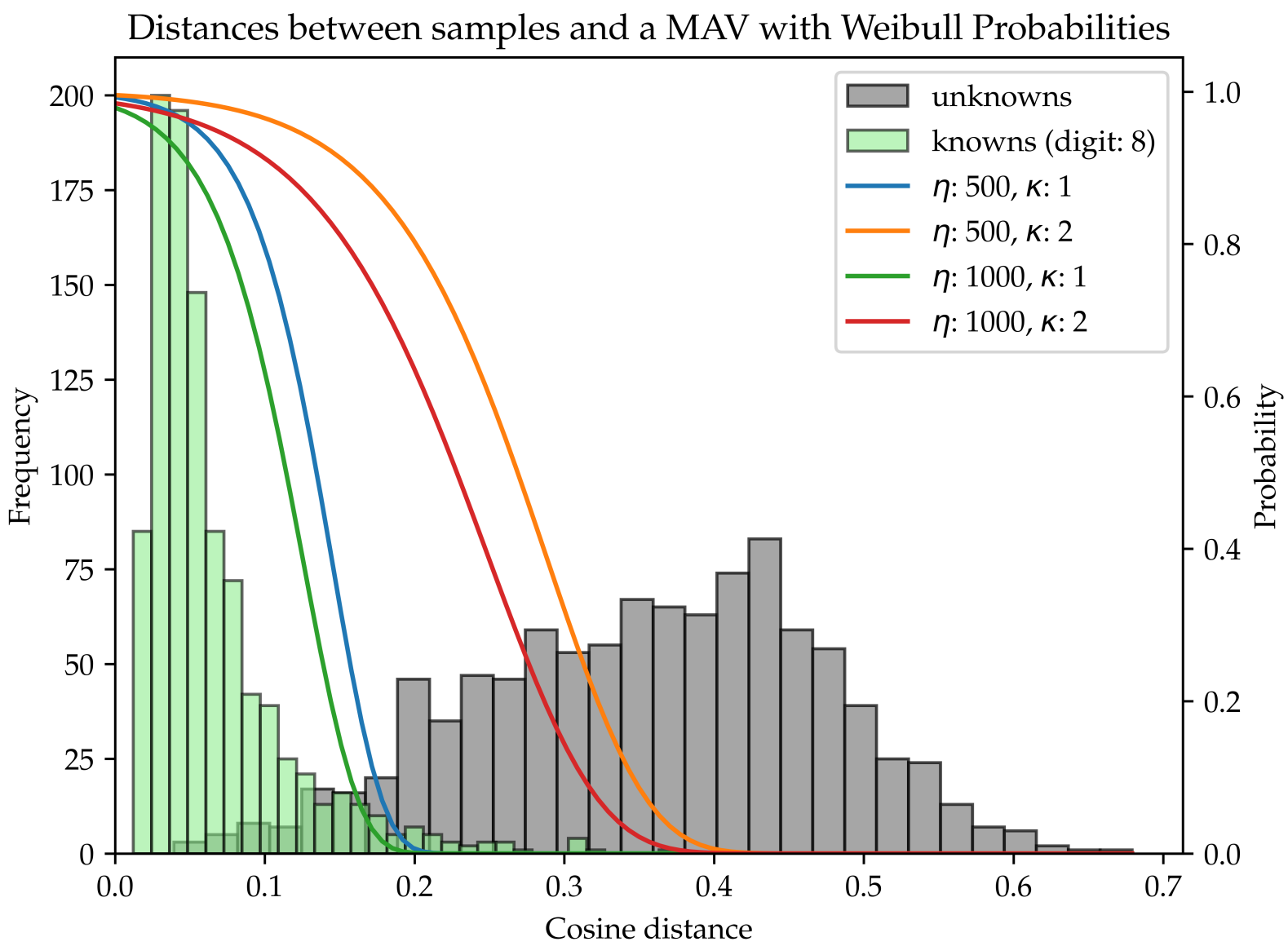
Calibrating OpenMax

- Training
- Class representation
- Mean Activation Vector μ
- Features from penultimate layer Γ
- Correctly classified samples
- Weibull Distribution
- Tail size η & Distance Multiplier κ

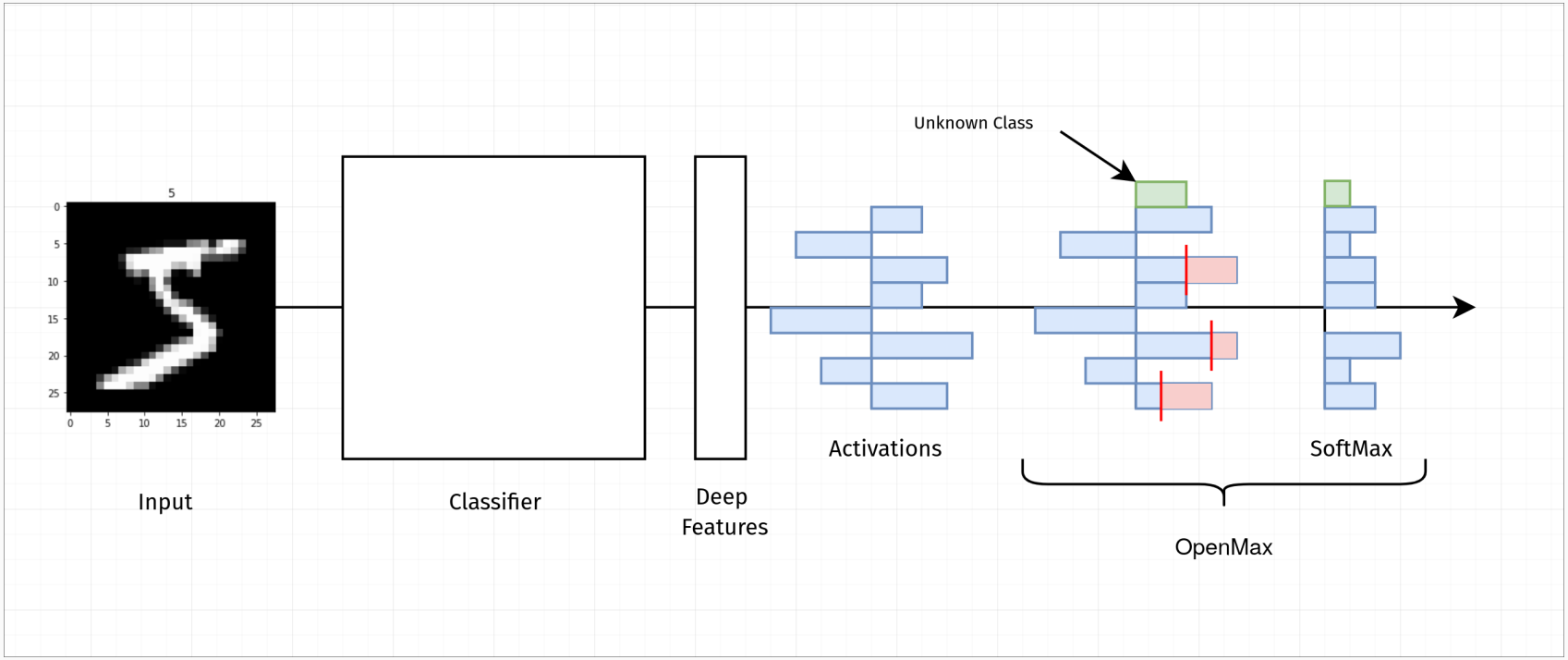
Projection of MAV



Building a Weibull Distribution



OpenMax Overview



OpenMax in Detail

For a testing sample:

1. Sort and select α classes (logit value)
2. Weibull Probability ω based on distance μ_i and Γ
3. $\hat{z} = z \circ \omega$
4. $\hat{z}_{N+1} = \sum_i z_i(1 - \omega_i)$
5. $\mathsf{softmax}(\hat{z})$

Approach & Experiments

Improving OpenMax

- Handling negative logits
  - Value Shift
  - Adjust Probabilities
- Introducing weight factors
  - $\phi_N = \frac{1}{N-1}$
  - $\phi_\omega = \frac{1}{\sum_i (1-\omega_i)}$
- Removing alpha parameter

Research Questions 1

RQ1: Can OpenMax performance be enhanced by accounting for negative activation values?

Open-Set Classification Rate

Correct Classification Rate (CCR)

- Known samples
- Threshold θ

$$\mathrm{CCR}(\theta) = \frac{|\{k_c | \operatorname{argmax}_{1 \leq n \leq N} y_{c,n} = \tau_c \wedge y_{c,n} \geq \theta\}|}{|K|} \quad (1)$$

False Positive Rate (FPR)

- Unknown samples
- Threshold θ

$$\mathrm{FPR}(\theta) = \frac{|\{u_c | \operatorname{argmax}_{1 \leq n \leq N} y_{c,n} \geq \theta\}|}{|U|} \quad (2)$$

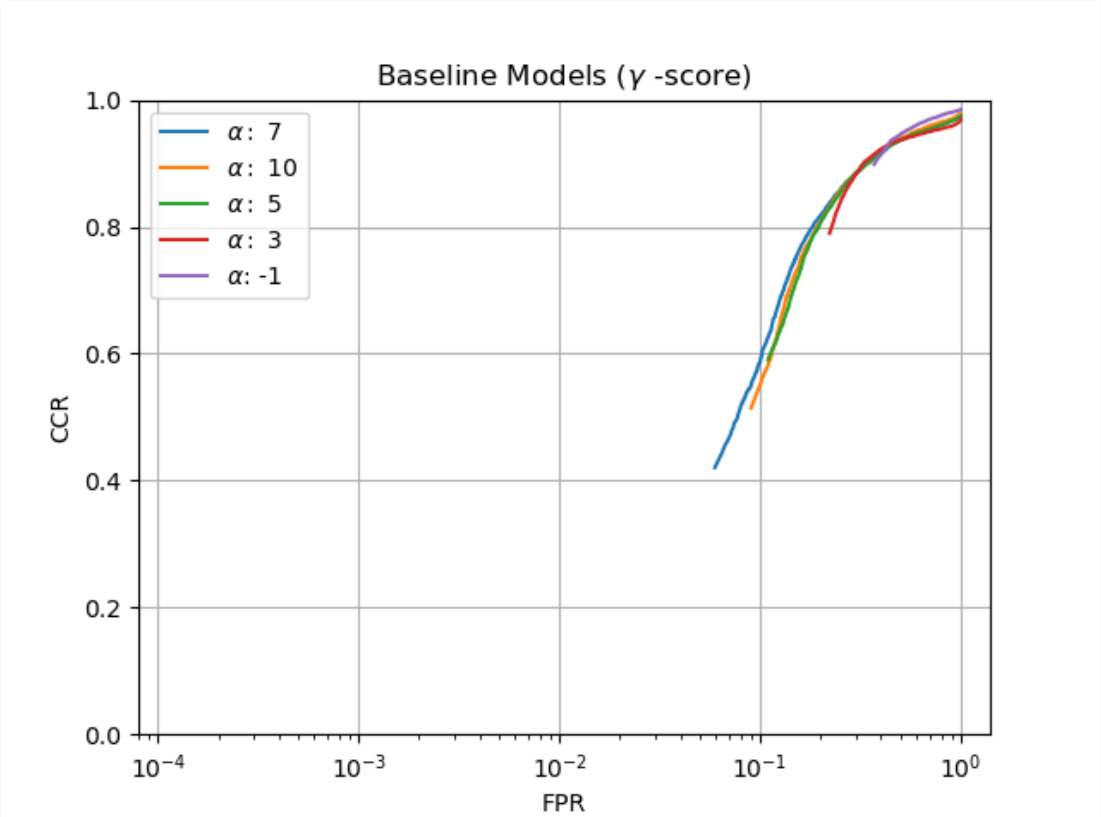
Defining Scoring System

- Σ-Score - Thresholding
- γ-Score $\gamma = \frac{(\gamma^+ + \gamma^-)}{2}$

$$\gamma^+ = \frac{1}{|K|} \sum_{c=1}^{|K|} y_{\tau_c} \tag{3}$$

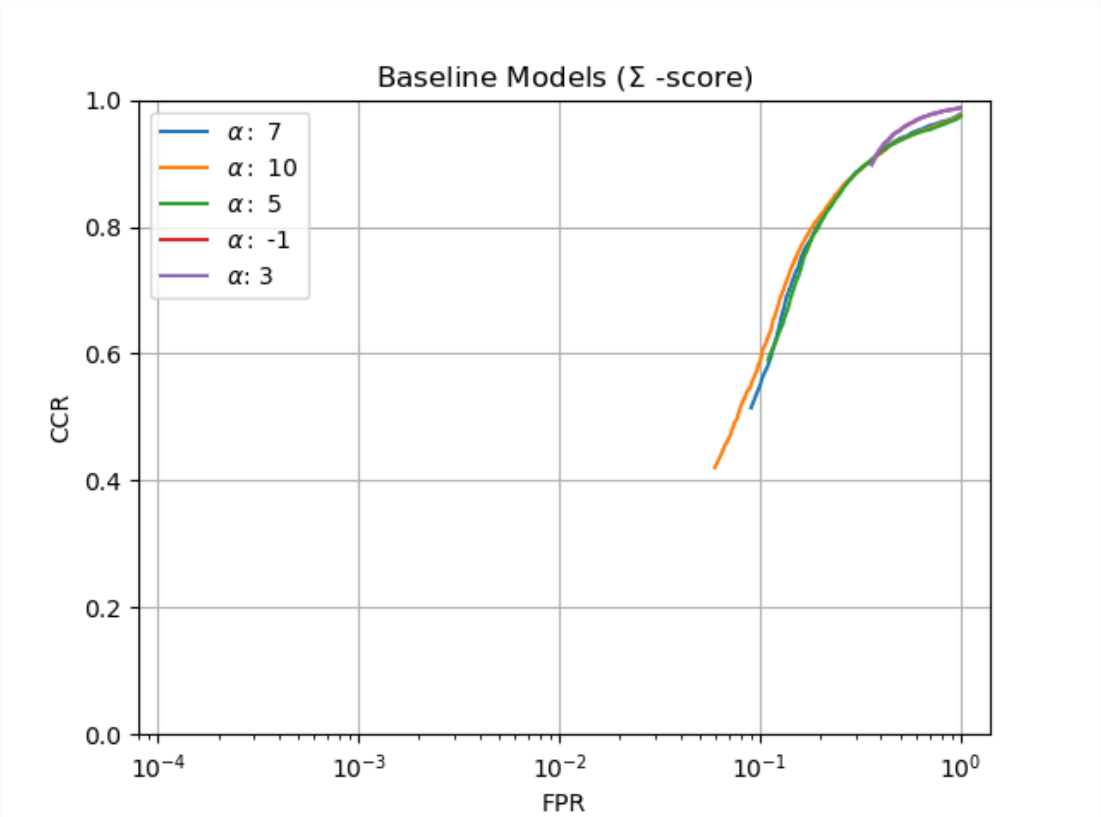
$$\gamma^- = \frac{1}{|U|} \sum_{c=1}^{|U|} \left(1 - \text{argmax}_{1 \leq n \leq N} y_{c,n}\right) \tag{4}$$

Baseline (γ)



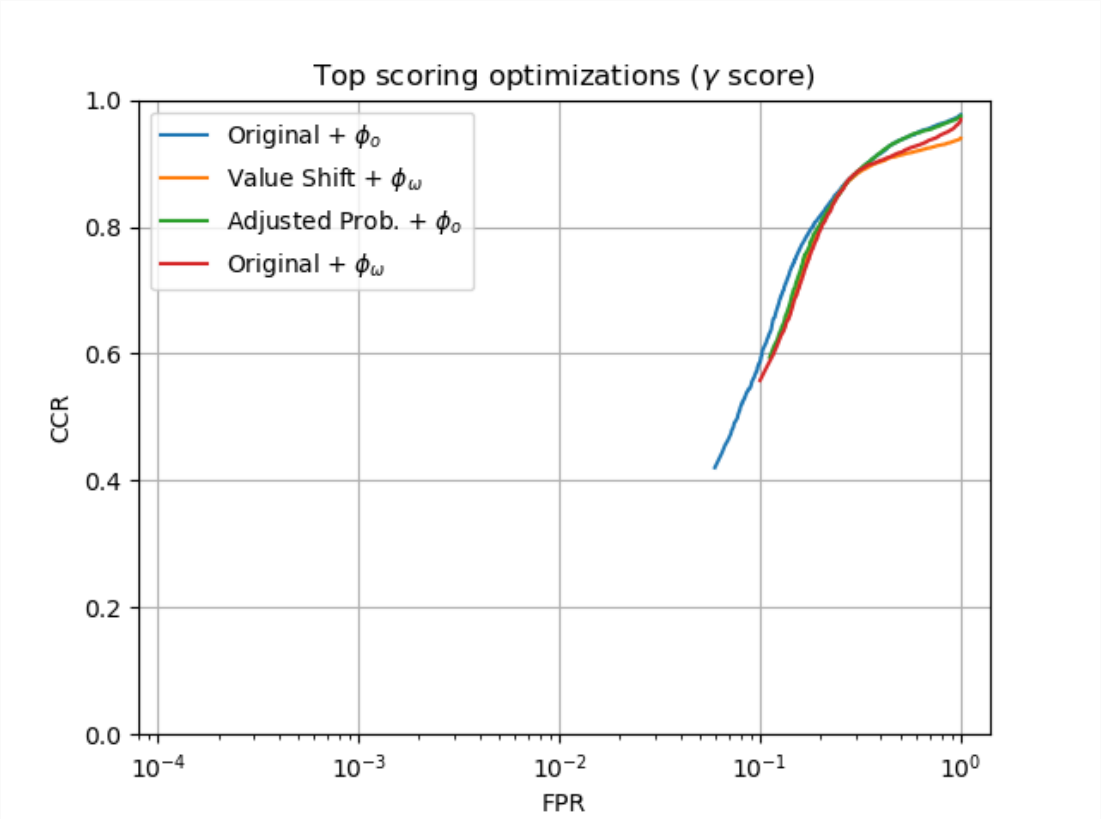
Rank	α	η	κ	γ	Σ
1	7	1000	1.0	0.805	4.272
2	10	1000	1.0	0.803	3.685
3	5	1000	1.0	0.800	3.669
4	3	1000	1.0	0.718	2.855
5	-1	1000	1.0	0.600	2.899

Baseline (Σ)



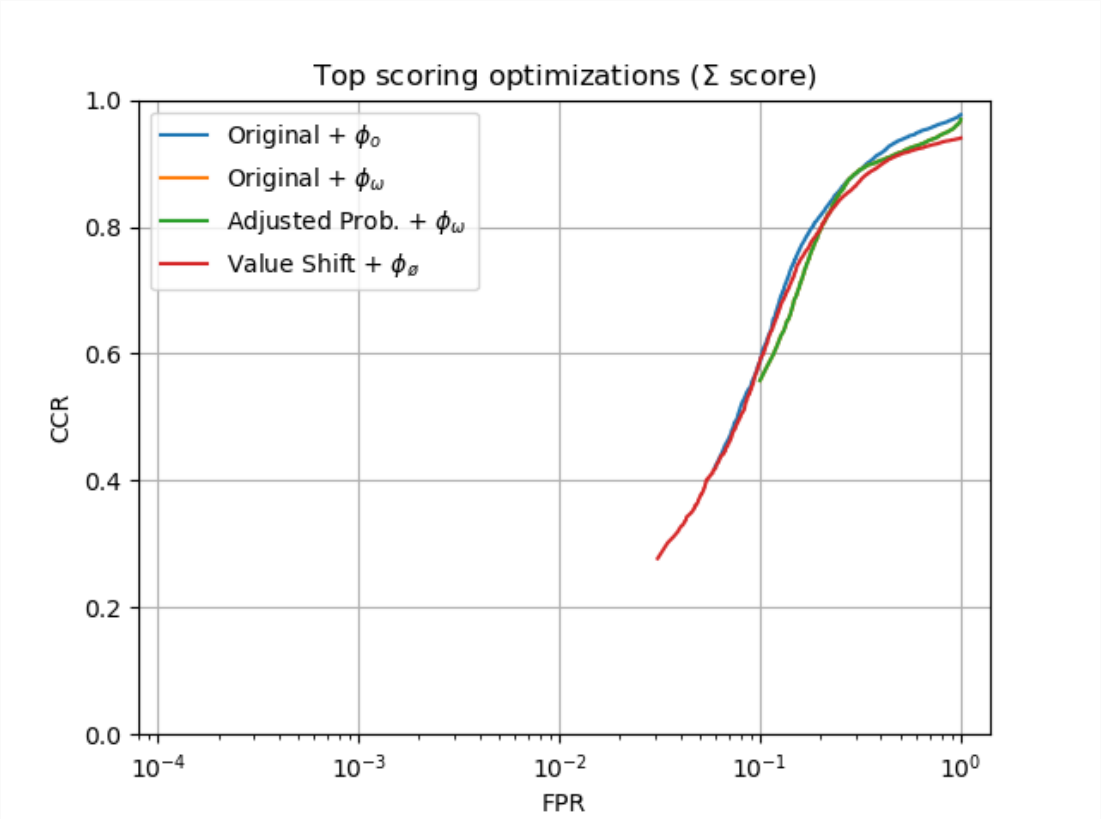
Rank	α	η	κ	γ	Σ
1	7	1000	1.0	0.805	4.272
2	10	1000	1.0	0.803	3.685
3	5	1000	1.0	0.800	3.669
4	-1	10	2.3	0.514	2.917
5	3	100	3.0	0.514	2.916

Results RQ1 (γ)



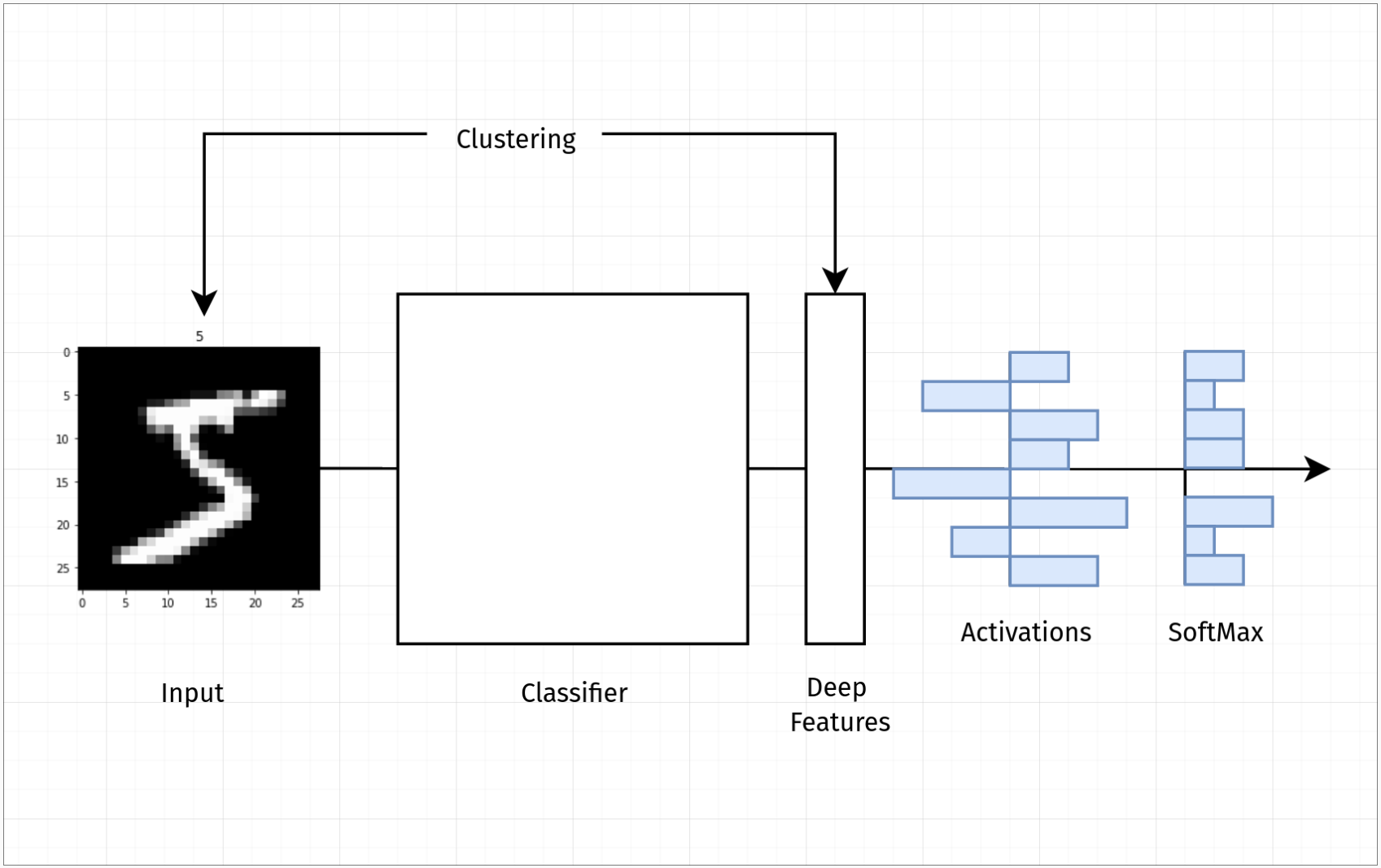
Rank	α	η	κ	ϕ	Negative fix	γ	Σ
1	7	1000	1.0	ϕ_o	Original	0.805	4.272
2	5	1000	1.0	ϕ_ω	Value Shift	0.801	3.578
3	5	1000	1.0	ϕ_o	Adjusted Prob.	0.801	3.671
4	3	1000	1.0	ϕ_ω	Original	0.792	4.176

Results RQ1 (Σ)



Rank	α	η	κ	ϕ	Negative fix	γ	Σ
1	7	1000	1.0	ϕ_o	Original	0.805	4.272
2	3	1000	1.0	ϕ_ω	Original	0.792	4.176
3	3	1000	1.0	ϕ_ω	Adjusted Prob.	0.792	4.176
4	5	1000	1.0	ϕ_o	Value Shift	0.749	4.169

Where to apply the clustering?



Introducing Clustering to OpenMax

Input Clustering:

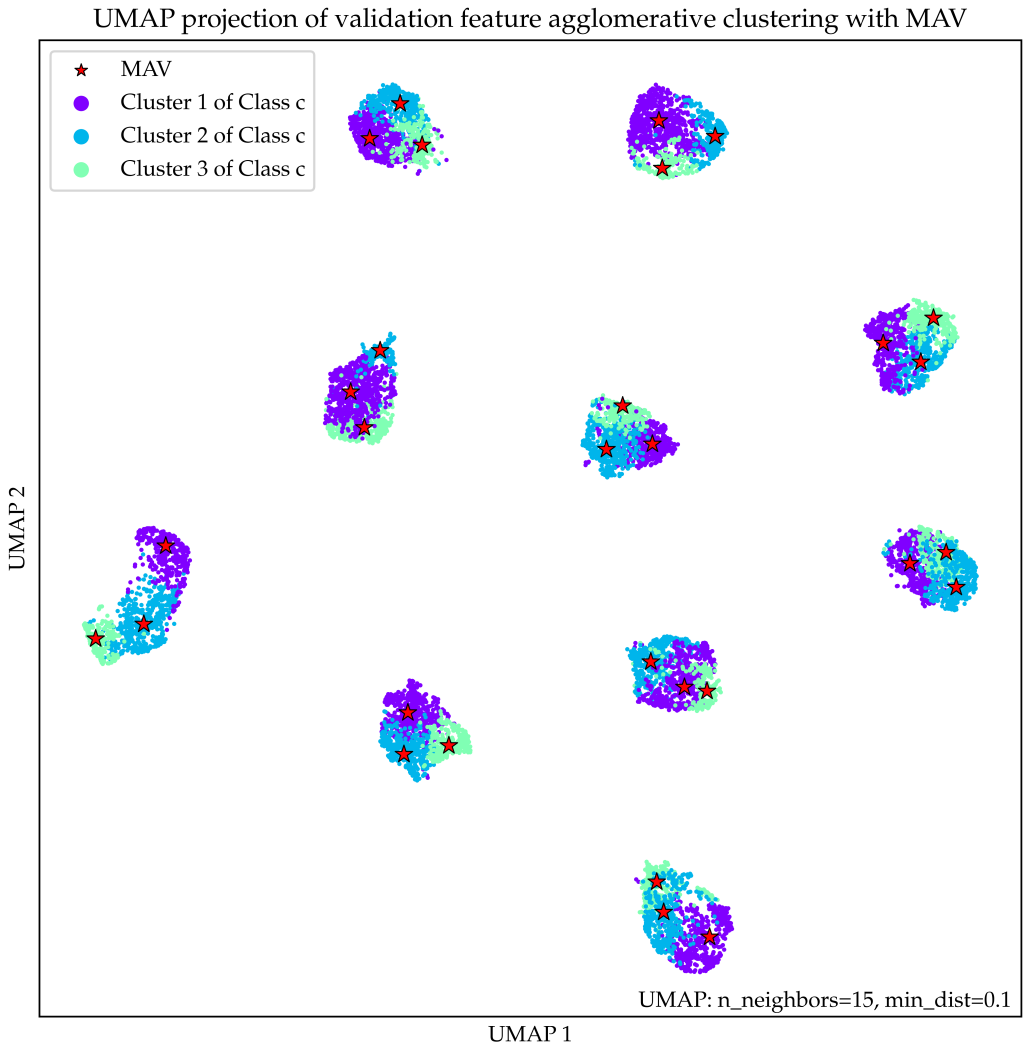
- Input data
- Per class
- Each cluster a class

Features Clustering:

- Training Features
- Validation Features
- Per class

Combination of both types

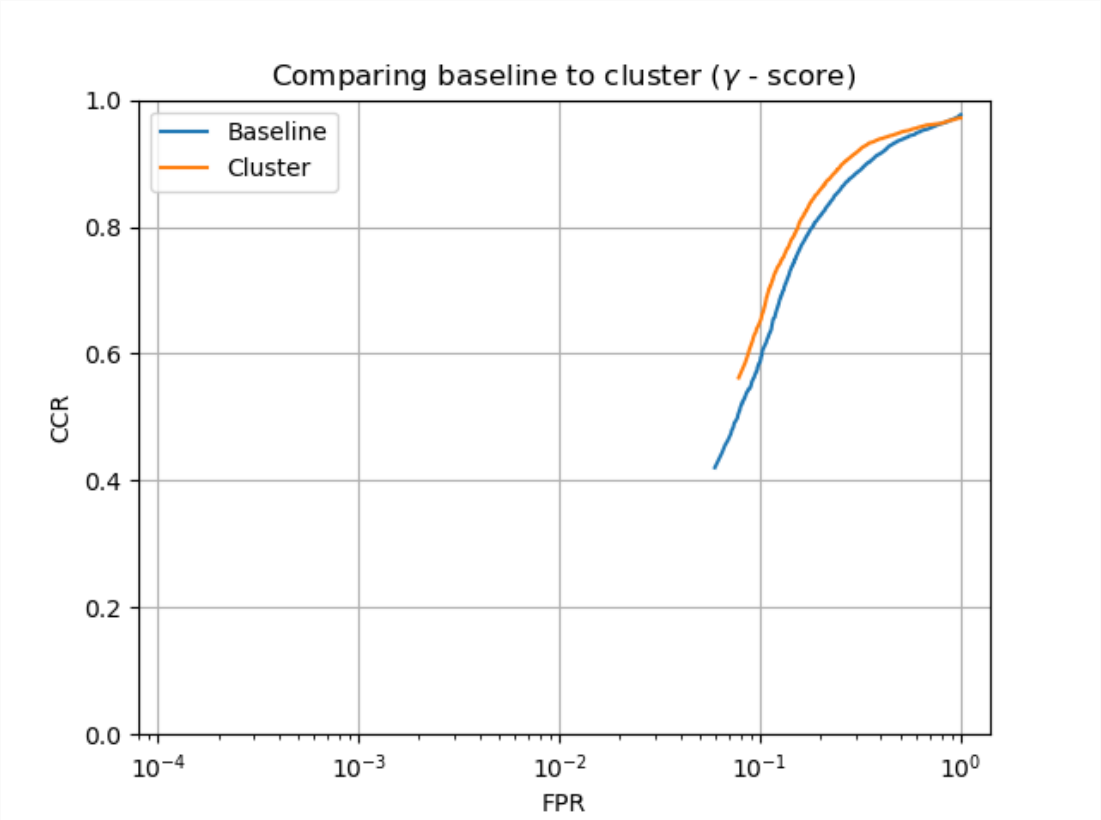
Visualizing Clustering



Research Questions 2

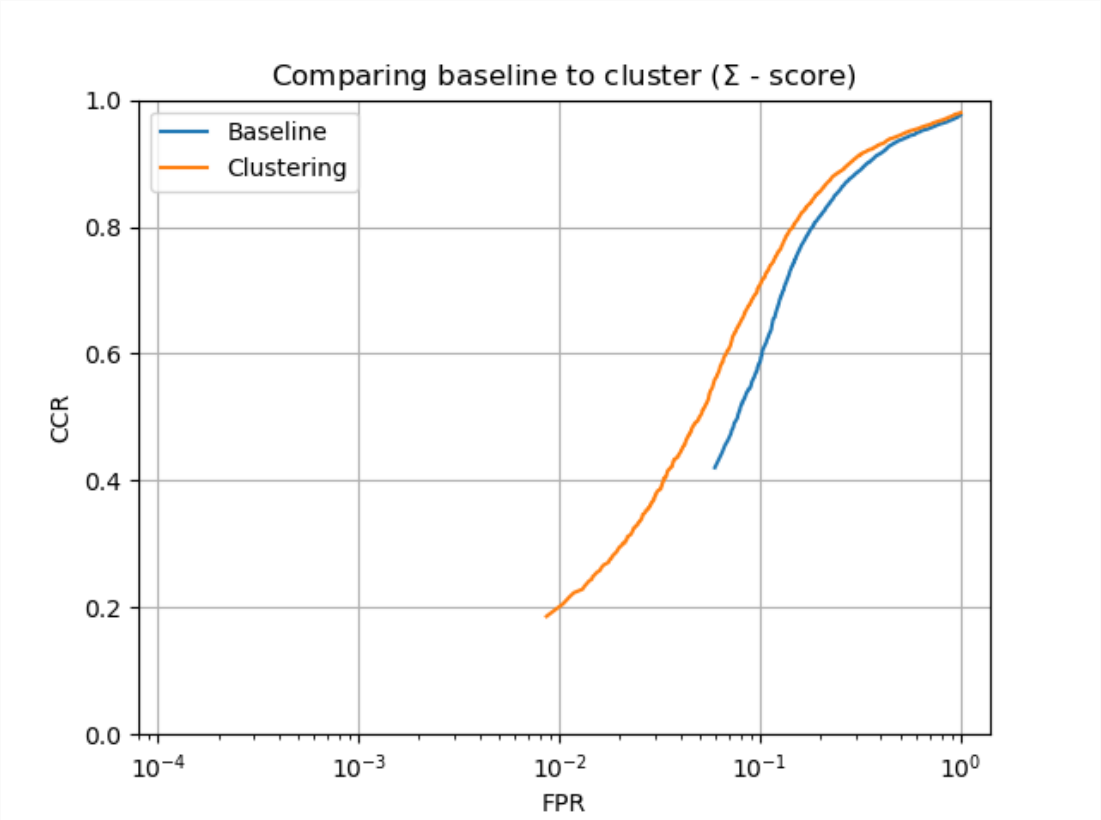
RQ2: Can clustering improve OpenMax's performance?

Results RQ2 (γ)



Rank	α	η	κ	k_{Input}	k_{Feat}	γ	Σ
1	7	100	0.7	1	5	0.829	4.391
2	7	1000	1.0	1	1	0.805	4.272

Results RQ2 (Σ)

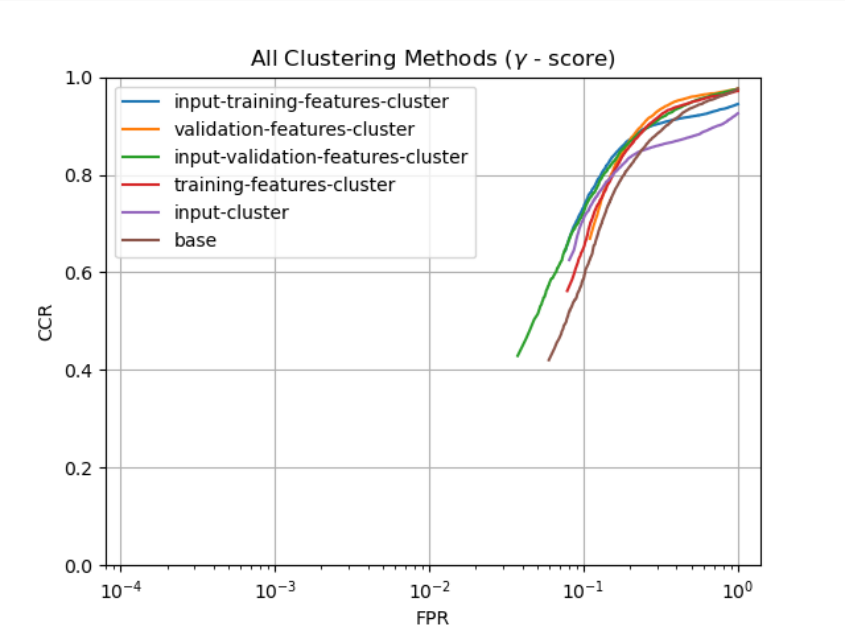


Rank	α	η	κ	k_{Input}	k_{Feat}	γ	Σ
1	7	750	1.25	1	6	0.768	4.450
2	7	1000	1.0	1	1	0.805	4.272

Research Questions 3

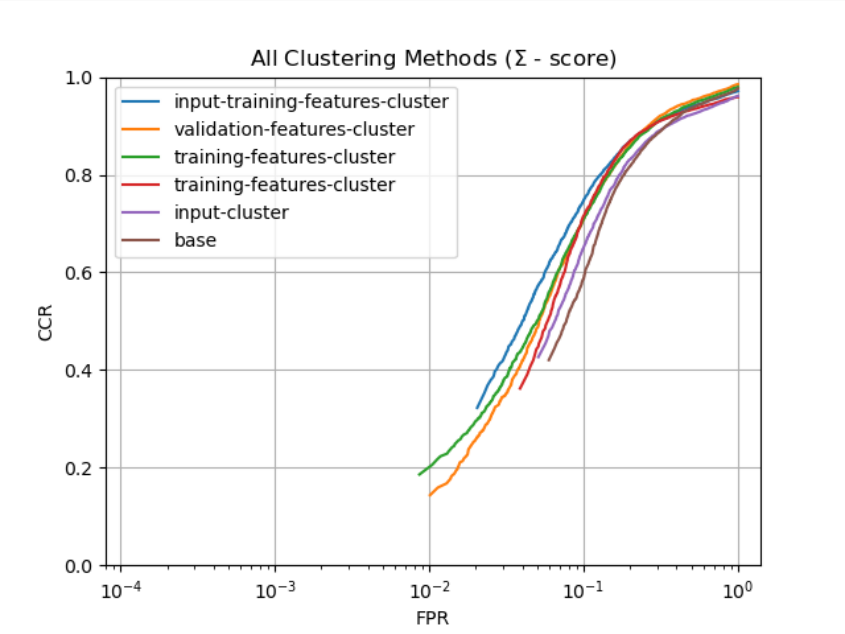
RQ3: If clustering improves OpenMax, which clustering type is optimal, and by using which parameters?

Results RQ3 (γ)



Rank	Type	α	η	κ	k_{Input}	k_{Feat}	γ	Σ
1	ITFC	5	100	0.7	2	3	0.841	3.666
2	VFC	7	10	0.5	1	5	0.834	3.774
3	IVFC	5	100	1.25	2	6	0.834	4.478
4	TFC	7	100	0.7	1	5	0.829	4.391
5	IC	5	100	0.5	2	1	0.822	4.247
6	base	7	1000	1.0	1	1	0.805	4.272

Results RQ3 (Σ)



Rank	Type	α	η	κ	k_{Input}	k_{Feat}	γ	Σ
1	IVFC	5	100	1.0	2	6	0.828	4.483
2	VFC	7	250	1.25	1	5	0.742	4.475
3	TFC	7	750	1.25	1	6	0.768	4.450
4	ITFC	5	1000	1.5	2	3	0.827	4.423
5	IC	7	500	1.0	2	1	0.812	4.307
6	base	7	1000	1.0	1	1	0.805	4.272

Discussion

Limitations & Future Work

- Datasets
- Other Clusetring Algorithms
- Per clusters parameters
- Costum Loss function

Questions?