BOREALIS AI

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

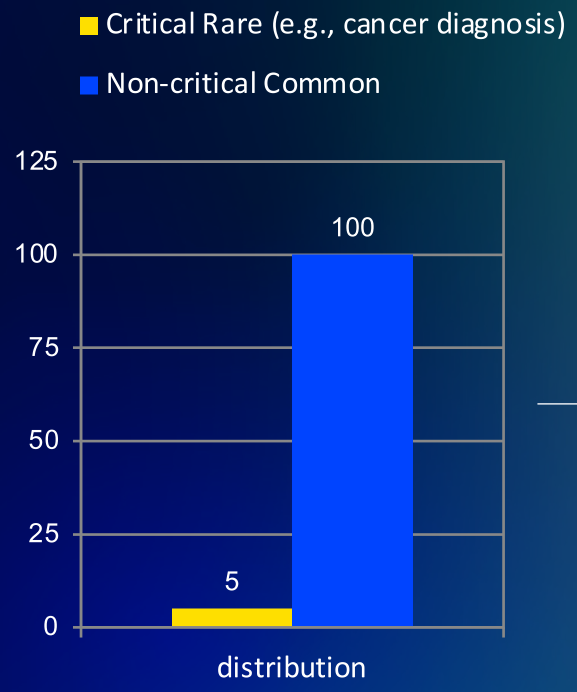
Ranking Regularization for Critical Rare Classes: Minimizing False Positives at a High True Positive Rate

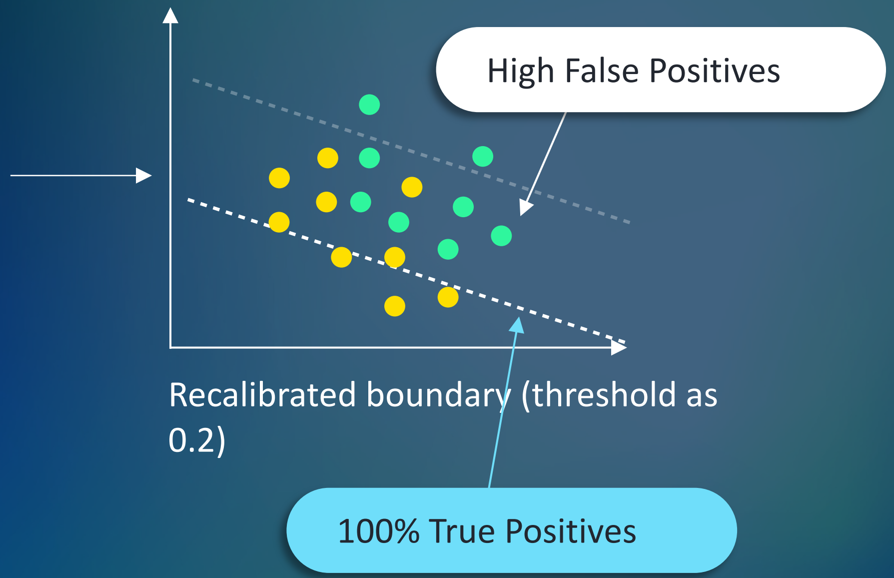
Kiarash Mohammadi [1,2], He Zhao[1], Mengyao Zhai[1], Frederick Tung[1]

1. Borealis AI
2. University of Montreal

Problem

- Imbalanced distribution
- Missing the critical class has a disproportionately high cost, e.g, Cancer diagnosis or fraud detection
- Current solution obtains high TPR at the cost of high FPR.

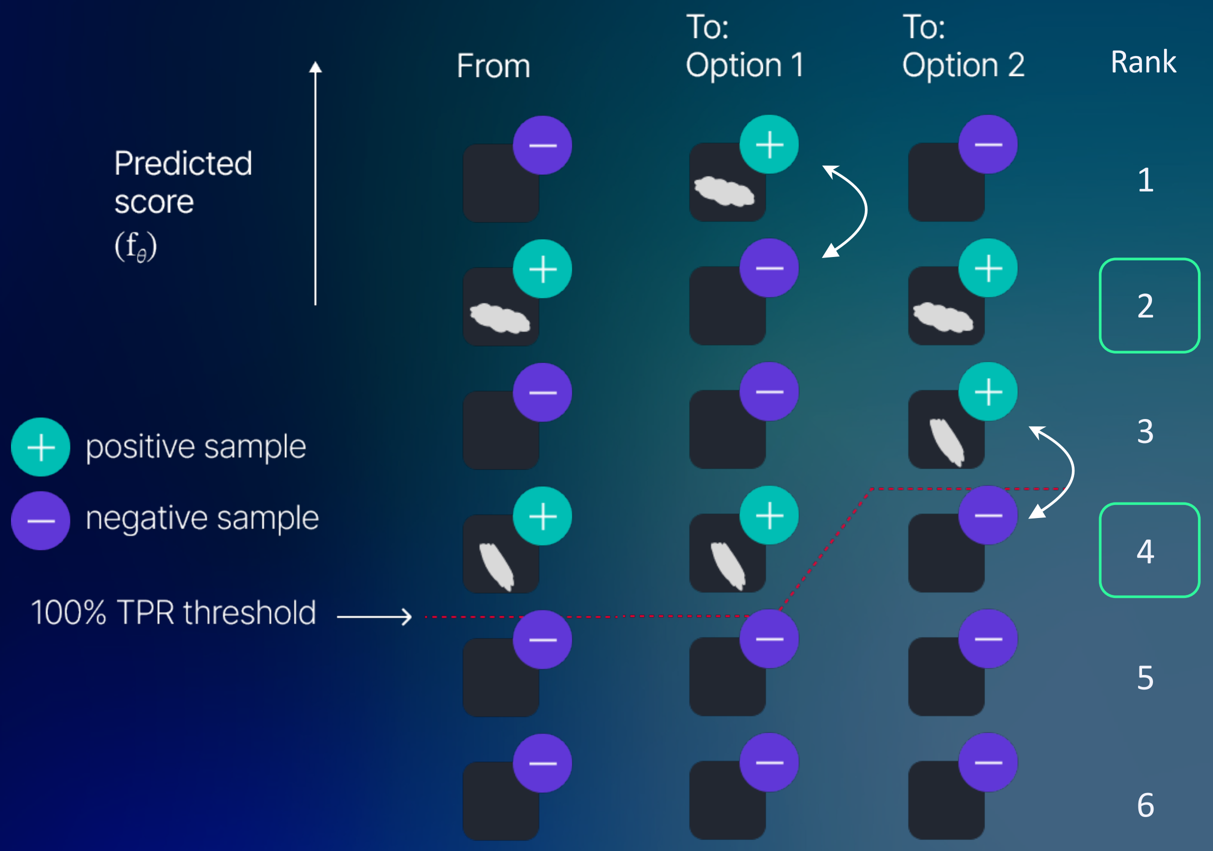




Proposed Regularization Method

→ The regularization term is then defined as:

$$R(\theta) = \frac{1}{|P|} \sum_{i=1}^{M} r_i^2 \cdot \mathbb{1}[y_i = 1]$$



	From	To: Option 1	To: Option 2
AUC (↑)	0.625	0.75	0.75
FPR@100% TPR (↓)	50%	50%	25%
ℓ_{reg}	20	17	13

Experiment results

Methods	Binary CIFAR10, imb. **1:100** FPR@98%TPR ↓	FPR@95%TPR ↓	FPR@92%TPR ↓	AUC ↑
BCE	56.0	45.0	29.0	91.2
+ALM	52.0	34.0	21.0	93.1
+RankReg	**47.1**	**26.2**	**20.6**	94.3
S-ML	59.0	40.0	26.0	91.7
+ALM	50.0	37.0	**24.0**	92.5
+RankReg	**45.6**	**31.4**	29.7	93.9
S-FL	59.0	40.0	27.0	91.7
+ALM	55.0	39.0	25.0	91.5
+RankReg	**53.3**	**35.4**	**20.7**	92.8
A-ML	54.0	36.0	23.0	92.4
+ALM	**45.0**	35.0	23.0	92.8
+RankReg	47.8	**28.9**	**21.4**	94.1
A-FL	50.0	38.0	24.0	92.3
+ALM	**49.0**	37.0	23.0	92.8
+RankReg	50.5	**28.7**	**20.9**	94.3
CB-BCE	89.0	72.0	59.0	78.0
+ALM	67.0	51.0	36.0	88.1
+RankReg	**48.8**	**29.9**	**24.6**	93.2
W-BCE	69.0	52.0	37.0	87.4
+ALM	66.0	48.0	31.0	89.3
+RankReg	**60.0**	**39.4**	**29.6**	92.1
LDAM	65.0	48.0	34.0	89.0
+ALM	60.0	42.0	31.0	91.0
+RankReg	**42.8**	**25.6**	**23.8**	95.0
Avg. Δ	6.0	9.7	2.8	2.3

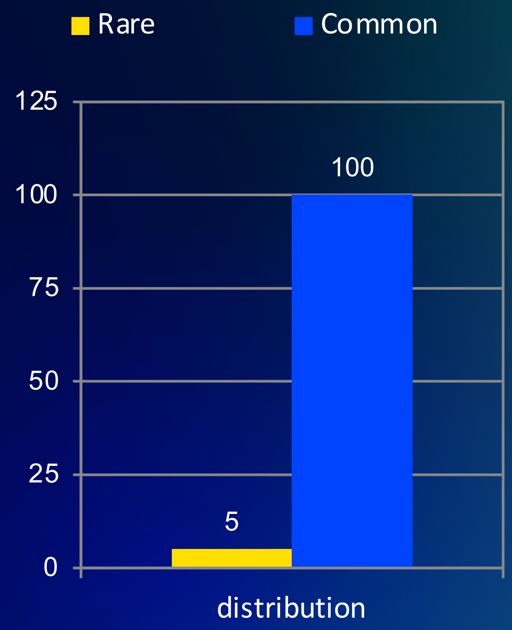
Methods	Melanoma, imb. **1:170** FPR@98%TPR ↓	FPR@95%TPR ↓	FPR@92%TPR ↓	FPR@90%TPR ↓	AUC ↑
BCE	49.8	45.9	38.6	35.5	85.7
+ALM	49.9	41.8	40.0	37.7	85.6
+RankReg	**49.4**	**37.9**	**33.9**	**31.6**	86.8
S-ML	**46.6**	42.8	38.4	37.4	85.3
+ALM	51.3	**40.5**	39.8	36.2	83.5
+RankReg	54.6	42.4	**36.1**	**34.4**	86.3
S-FL	59.0	47.3	44.4	39.5	83.8
+ALM	**47.8**	42.7	39.2	38.1	84.0
+RankReg	56.6	**37.8**	**31.2**	**29.8**	86.1
A-ML	**47.5**	42.9	40.4	36.6	85.4
+ALM	51.0	41.5	37.5	37.1	83.7
+RankReg	58.3	**40.8**	**36.7**	**33.9**	86.2
A-FL	55.6	45.0	42.7	41.2	84.4
+ALM	49.0	42.4	40.1	38.1	83.6
+RankReg	**48.0**	**36.2**	**30.7**	**28.8**	86.3
W-BCE	69.0	52.0	37.0	32.1	87.4
+ALM	66.0	48.0	**31.0**	30.7	89.3
+RankReg	**56.4**	**41.1**	33.0	**30.5**	90.9
LDAM	**59.7**	48.2	46.2	**39.0**	83.4
+ALM	62.7	47.7	**43.3**	40.7	81.5
+RankReg	65.6	**47.5**	45.7	43.9	81.7

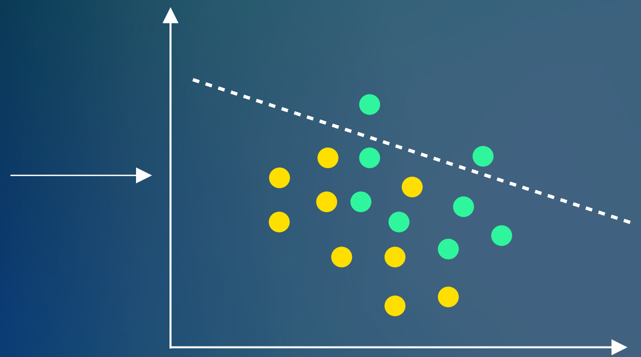
Outline

- → Problem and Motivation
- → Proposed ranking regularization
- → Experiment details and Results

Problem

→ Imbalanced distribution

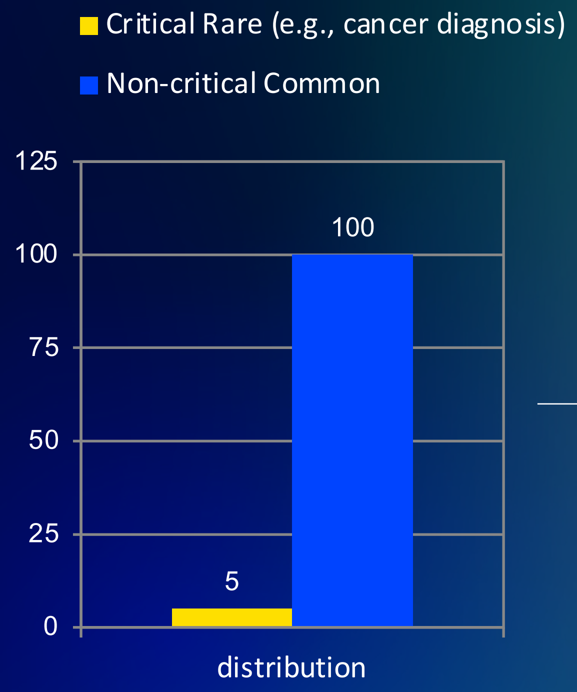


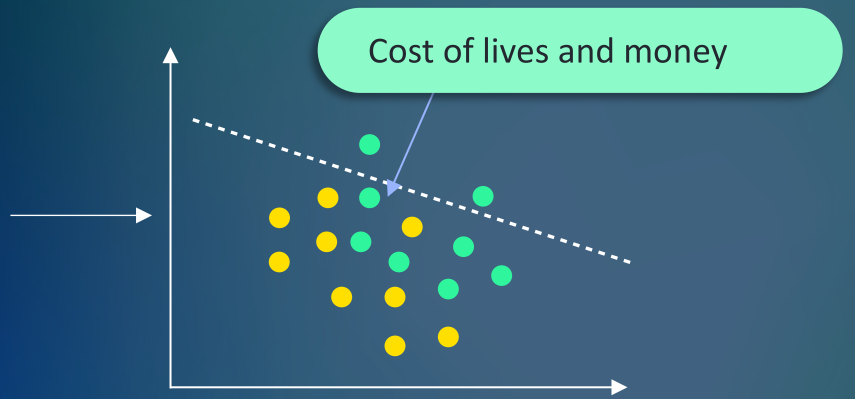


Biased decision boundary (threshold as 0.5).

Problem

- Imbalanced distribution
- Missing the critical class has a disproportionately high cost, e.g, Cancer diagnosis or fraud detection

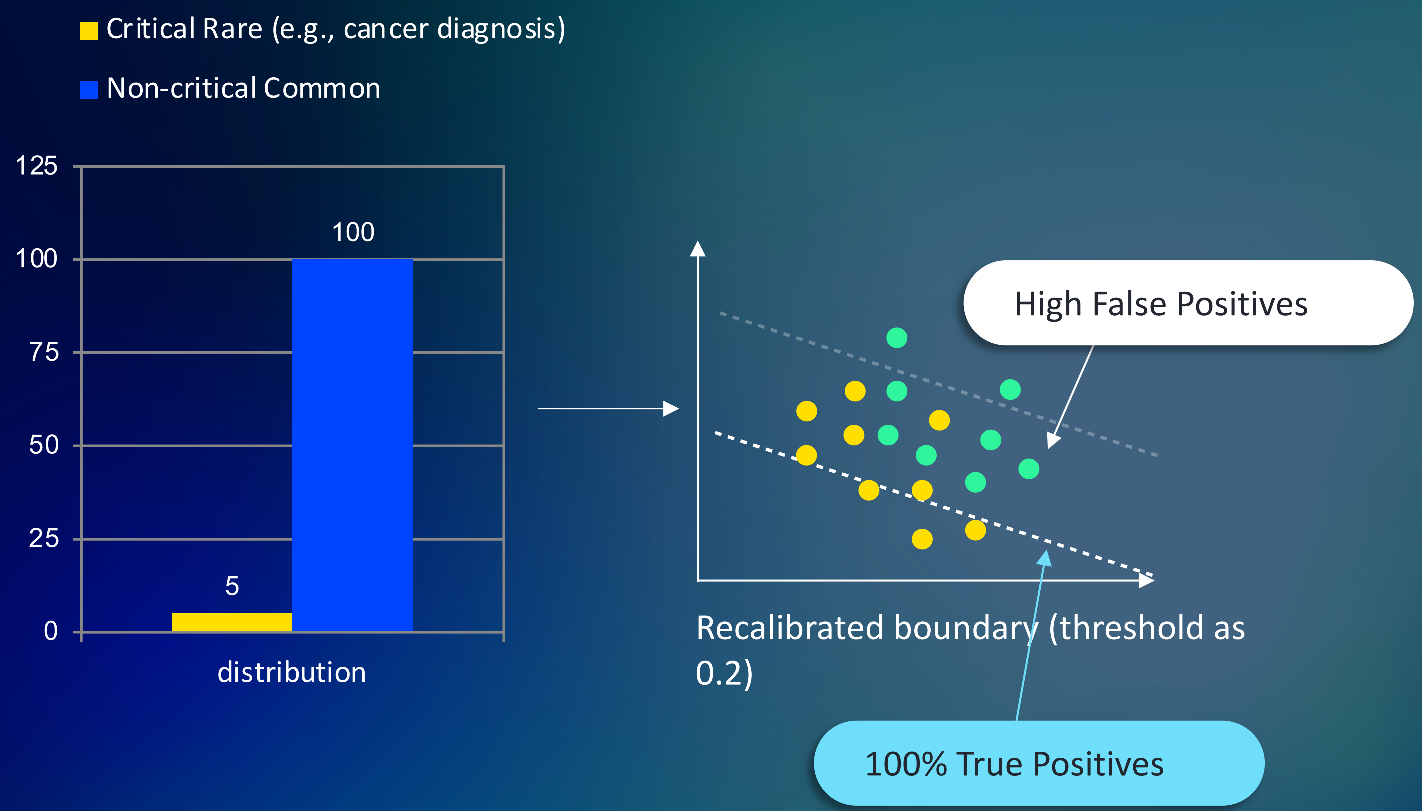




Biased decision boundary (threshold as 0.5).

Problem

- Imbalanced distribution
- Missing the critical class has a disproportionately high cost, e.g, Cancer diagnosis or fraud detection
- Current solution obtains high TPR at the cost of high FPR.



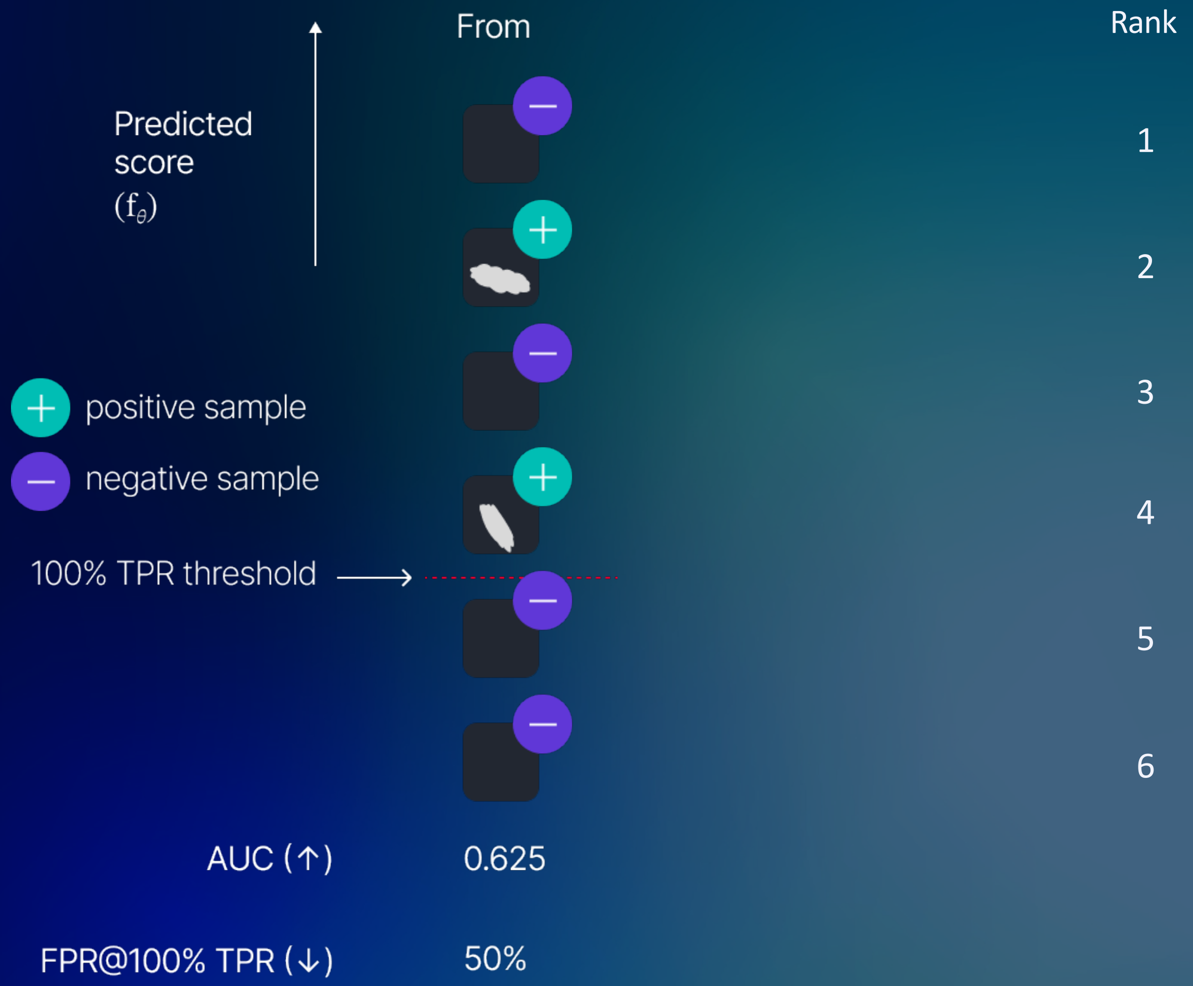
BOREALIS AI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

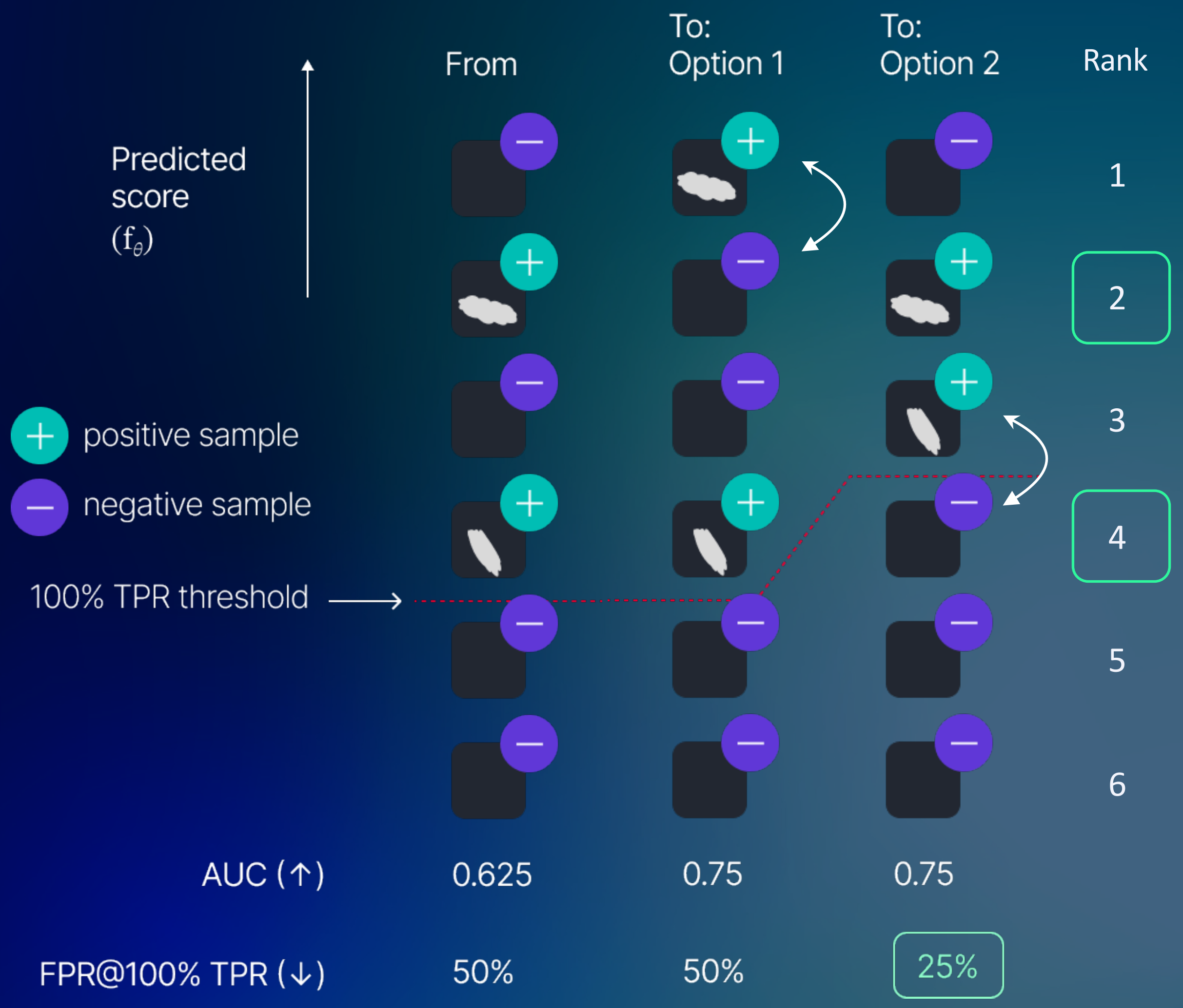
Outline

- → Problem and Motivation
- → Proposed ranking regularization
- → Experiment details and Results

Proposed Regularization Method



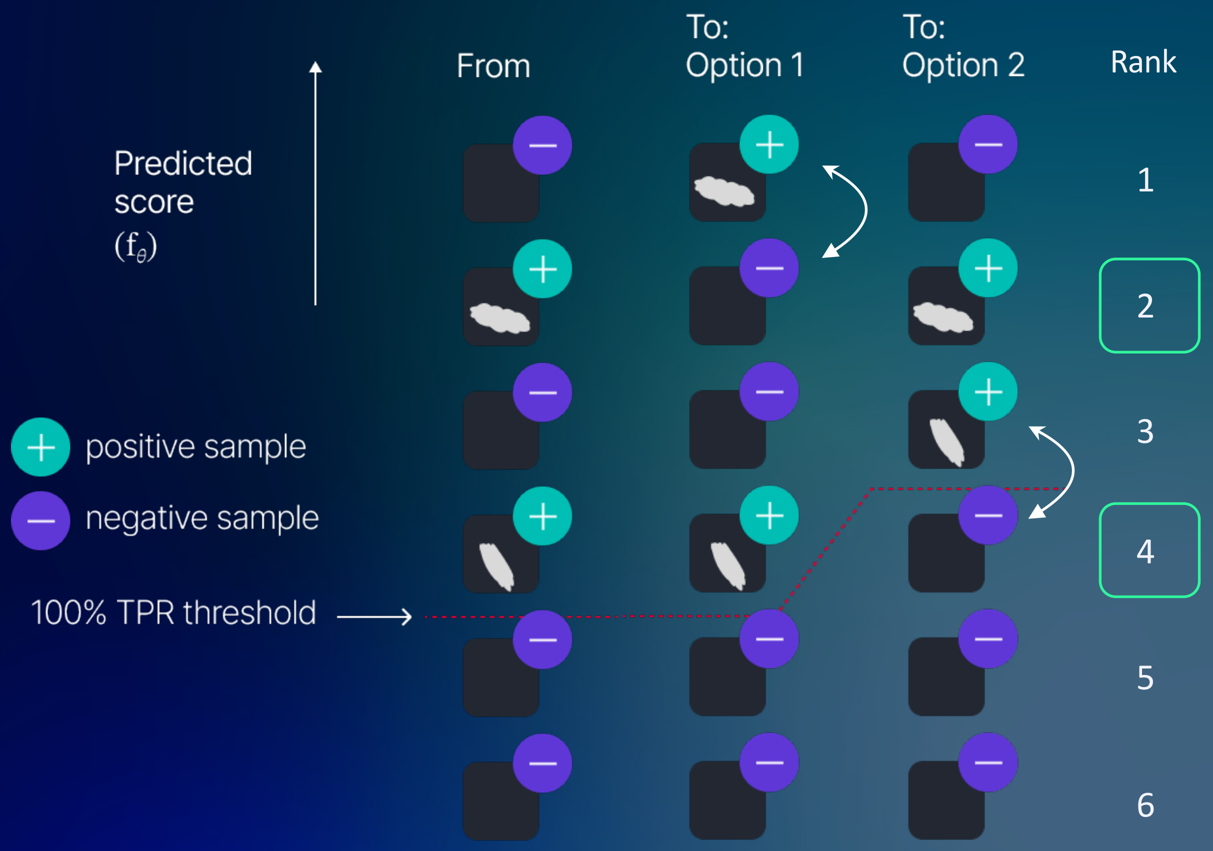
Proposed Regularization Method



Proposed Regularization Method

→ The regularization term is then defined as:

$$R(\theta) = \frac{1}{|P|}\sum_{i=1}^{M} r_i^2 \cdot \mathbb{1}[y_i = 1]$$



	From	To: Option 1	To: Option 2
AUC (↑)	0.625	0.75	0.75
FPR@100% TPR (↓)	50%	50%	25%
ℓ_{reg}	20	17	13

Proposed Regularization Method

There's a problem!

Ranking function gradients are zero almost everywhere

$$\mathbf{rk}(\mathbf{a})_j = 1 + |\{k : \mathbf{a}_k > \mathbf{a}_j\}|$$

$$\mathbf{r} = \mathbf{rk}([0.3, 0.1, 0.7, 0.6, 0.8]) = [4, 5, 2, 3, 1]$$

How to get meaningful gradients?

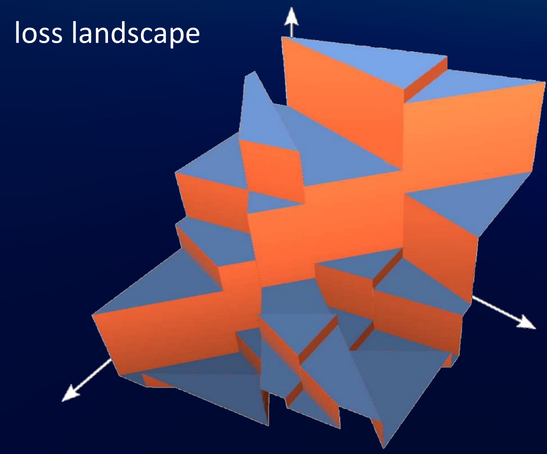
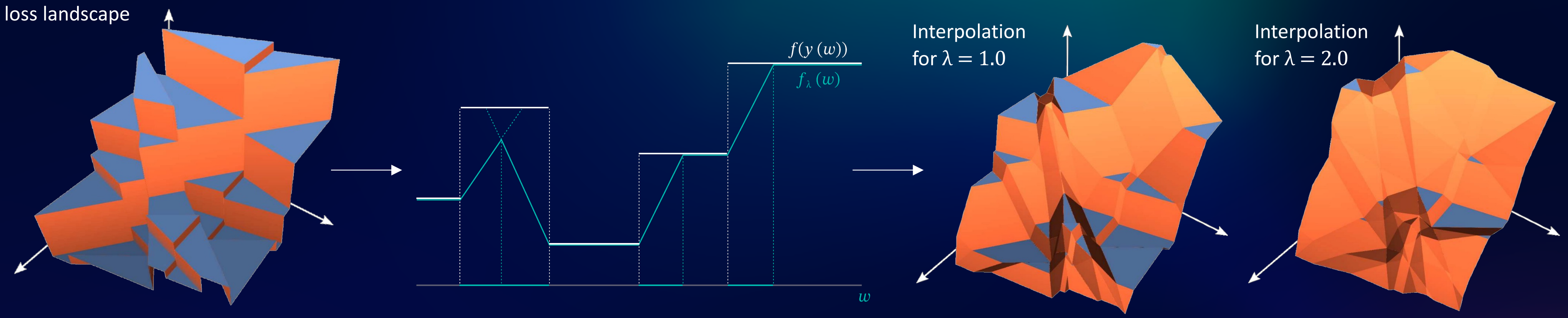


Figure from Michal, et al., *Optimizing Rank-based Metrics with Blackbox Differentiation*, CVPR'20

Blackbox Differentiation of Combinatorial Solvers



1. Figure from Marin, et al., *Differentiation of blackbox combinatorial solvers*, ICLR'20
2. Figure from Michal, et al., *Optimizing Rank-based Metrics with Blackbox Differentiation*, CVPR'20

BOREALIS AI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Outline

- → Problem and Motivation
- → Proposed ranking regularization
- → **Experiment details and Results**

Experiment details

Using *RankReg* with eight different base loss functions, as

→ Base loss varies from production to production.

→ Sensitivity test.

Baseline methods. Following previous method [24], we consider applying our proposed regularizer with several different existing loss functions, most of which have been designed to handle class imbalance: binary cross-entropy (BCE), symmetric margin loss (S-ML) [20], symmetric focal loss (S-FL) [18], asymmetric margin loss (A-ML) and focal loss (A-FL) [17], cost-weighted BCE (WBCE) [35], class-balanced BCE (CB-BCE) [8], and label distribution aware margin (LDAM) [4].

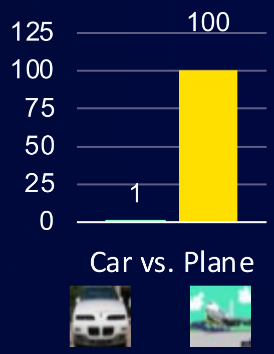
$$\arg\min_{\theta} F(\theta) + \alpha \cdot R(\theta)$$

RankReg

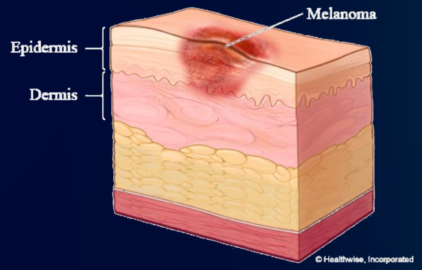
Experiment details: Datasets

- We compare against the state-of-the-art method, ALM*.
- Datasets:

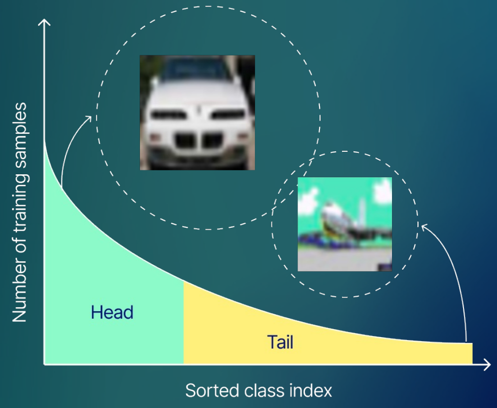
- Imbalanced CIFAR10/100 datasets
- Curated imbalanced version
- Ratios: 1 vs. 100 / 200



- Melanoma dataset
- Skin cancer classification
- Naturally imbalanced
- Ratio: 1 vs. 170



- Long-Tailed (LT) CIFAR
- Multi-class classification



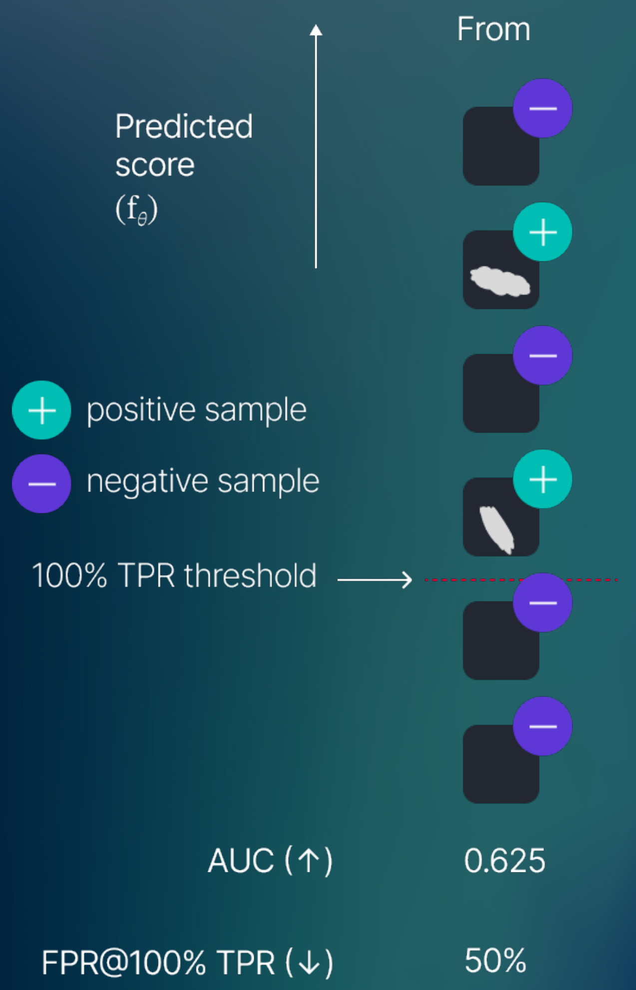
* Sangalli et al, Constrained Optimization to Train Neural Networks on Critical and Under-Represented Classes, NeurIPS21

BOREALIS AI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experiment details: Metrics

- We compare evaluate on the following metrics
- False Positives Rate at 98% True Positive Rate
- False Positives Rate at 95% True Positive Rate
- False Positives Rate at 92% True Positive Rate
- False Positives Rate at 90% True Positive Rate (Optional)
- Area Under ROC Curve (AUC)



Experiment results: CIFAR10 – Ratio 1 vs. 100

Binary CIFAR10, imb. **1:100**				
Methods	FPR@↓ 98%TPR	FPR@↓ 95%TPR	FPR@↓ 92%TPR	AUC↑
BCE	56.0	45.0	29.0	91.2
+ALM	52.0	34.0	21.0	93.1
+RankReg	**47.1**	**26.2**	**20.6**	94.3
S-ML	59.0	40.0	26.0	91.7
+ALM	50.0	37.0	**24.0**	92.5
+RankReg	**45.6**	**31.4**	29.7	93.9
S-FL	59.0	40.0	27.0	91.7
+ALM	55.0	39.0	25.0	91.5
+RankReg	**53.3**	**35.4**	**20.7**	92.8
A-ML	54.0	36.0	23.0	92.4
+ALM	**45.0**	35.0	23.0	92.8
+RankReg	47.8	**28.9**	**21.4**	94.1
A-FL	50.0	38.0	24.0	92.3
+ALM	**49.0**	37.0	23.0	92.8
+RankReg	50.5	**28.7**	**20.9**	94.3
CB-BCE	89.0	72.0	59.0	78.0
+ALM	67.0	51.0	36.0	88.1
+RankReg	**48.8**	**29.9**	**24.6**	93.2
W-BCE	69.0	52.0	37.0	87.4
+ALM	66.0	48.0	31.0	89.3
+RankReg	**60.0**	**39.4**	**29.6**	92.1
LDAM	65.0	48.0	34.0	89.0
+ALM	60.0	42.0	31.0	91.0
+RankReg	**42.8**	**25.6**	**23.8**	95.0
Avg. Δ	6.0	9.7	2.8	2.3

RankReg improves AUCs when coupling with base losses.

RankReg reduce SoTA FPRs by (2-9)% on various TPRs

BOREALIS AI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experiment results: Melanoma – Ratio 1 vs. 170

	Melanoma, imb. **1:170**				
Methods	FPR@ ↓ 98%TPR	FPR@ ↓ 95%TPR	FPR@ ↓ 92%TPR	FPR@ ↓ 90%TPR	AUC ↑
BCE	49.8	45.9	38.6	35.5	85.7
+ALM	49.9	41.8	40.0	37.7	85.6
+RankReg	**49.4**	**37.9**	**33.9**	**31.6**	86.8
S-ML	**46.6**	42.8	38.4	37.4	85.3
+ALM	51.3	**40.5**	39.8	36.2	83.5
+RankReg	54.6	42.4	**36.1**	**34.4**	86.3
S-FL	59.0	47.3	44.4	39.5	83.8
+ALM	**47.8**	42.7	39.2	38.1	84.0
+RankReg	56.6	**37.8**	**31.2**	**29.8**	86.1
A-ML	**47.5**	42.9	40.4	36.6	85.4
+ALM	51.0	41.5	37.5	37.1	83.7
+RankReg	58.3	**40.8**	**36.7**	**33.9**	86.2
A-FL	55.6	45.0	42.7	41.2	84.4
+ALM	49.0	42.4	40.1	38.1	83.6
+RankReg	**48.0**	**36.2**	**30.7**	**28.8**	86.3
W-BCE	69.0	52.0	37.0	32.1	87.4
+ALM	66.0	48.0	**31.0**	30.7	89.3
+RankReg	**56.4**	**41.1**	33.0	**30.5**	90.9
LDAM	**59.7**	48.2	46.2	**39.0**	83.4
+ALM	62.7	47.7	**43.3**	40.7	81.5
+RankReg	65.6	**47.5**	45.7	43.9	81.7

Extension on Multi-Class: Long-Tailed (LT) CIFAR

Error@β%TPR $\downarrow$	**LT-CIFAR10** imb. 100			**LT-CIFAR10** imb. 200		
	80%	90%	Acc.	80%	90%	Acc
CE	29.8	34.7	70.4	37.8	42.4	64.0
CE+ALM	28.9	33.9	70.9	36.1	39.9	65.1
CE+RankReg	26.7	29.3	71.6	36.7	37.8	65.0

Conclusions

- This paper introduces a general (plug-and-play) method to prioritize the reduction of false positives when the operational context calls for a high true positive rate.
- Ranking regularizer places an increasing penalty on positive samples the lower they are ranked in a sorted list of the network's classification scores.
- Experimental results show how our regularizer can be combined with a wide range of conventional losses and achieve state-of-the-art performance in standard metrics.