Multilabel reductions

What is my loss optimising?

Aditya Krishna Menon

Ankit Singh Rawat

Sashank Reddi

Sanjiv Kumar

Google Research

Multilabel classification

Given an instance, predict a binary label vector:

				
	1	1	0	0
	0	1	0	1

Multilabel predictions

Multilabel observations: $(x, \boldsymbol{y})$ where $\boldsymbol{y} \in \{0, 1\}^L$ # of labels

predictions: $f(x) \in \mathbb{R}^L$

Want to assign high scores to positive labels ($\boldsymbol{y}_i = 1$)

$\boldsymbol{y}$ =	0	1	1	0	1	0	1	0	1

$\boldsymbol{f}$ =	0.3	0.8	0.6	0.2	0.9	0.7	0.4	0.1	0.5

Google

Multilabel algorithm: one-versus-all

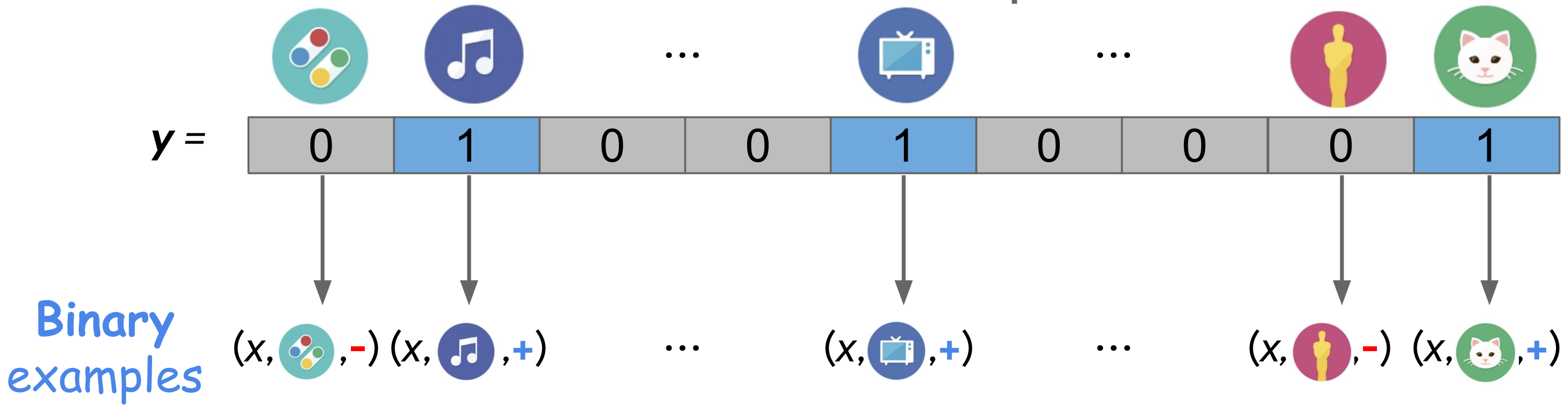
Create a binary-classification example for each label

y =

0	1	0	0	1	0	0	0	1

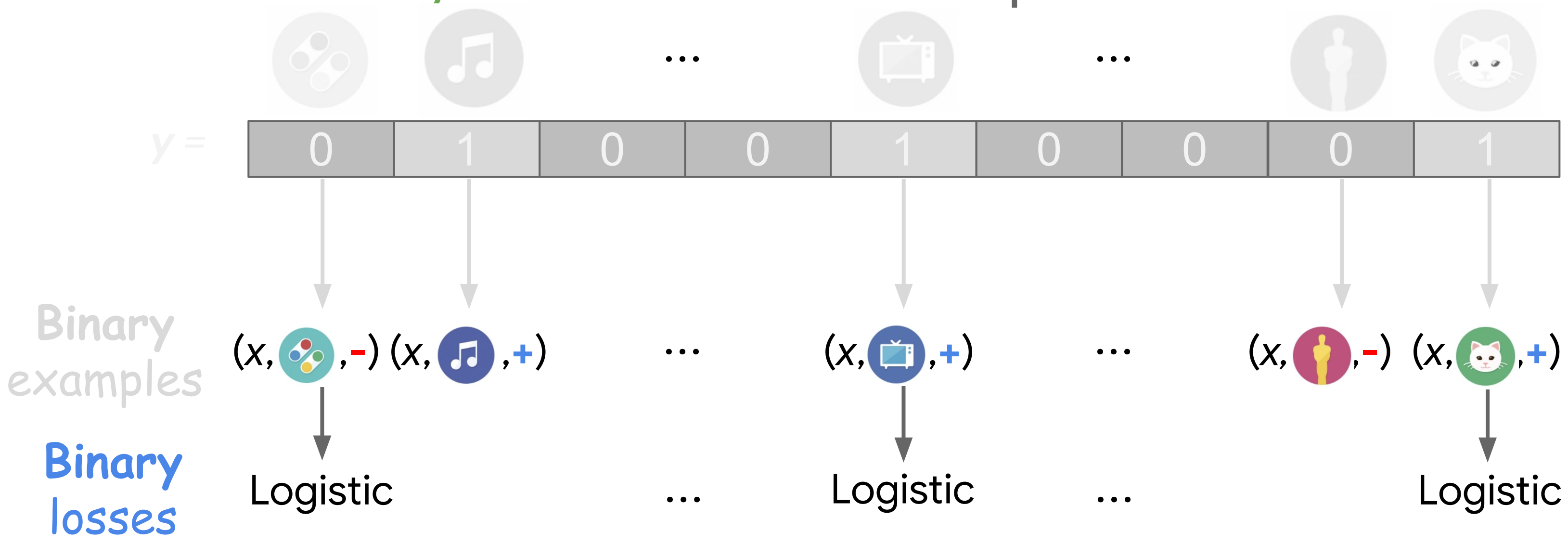
Multilabel algorithm: one-versus-all

Create a binary-classification example for each label



Multilabel algorithm: one-versus-all

Create a binary-classification example for each label



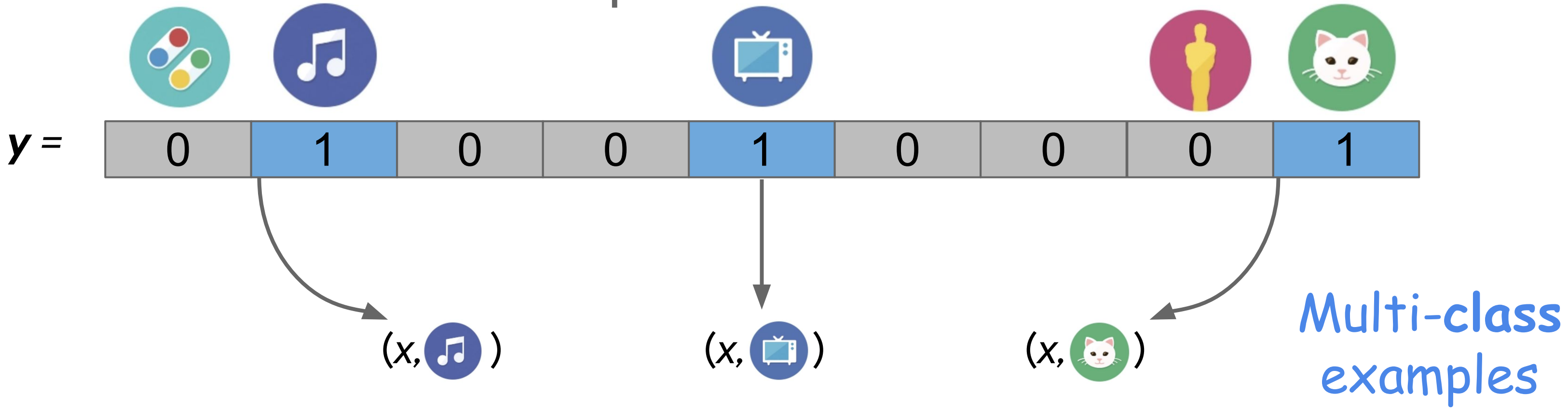
Multilabel algorithm: pick-all-labels

Create a multi-class example for each +'ve label

y =									
	0	1	0	0	1	0	0	0	1

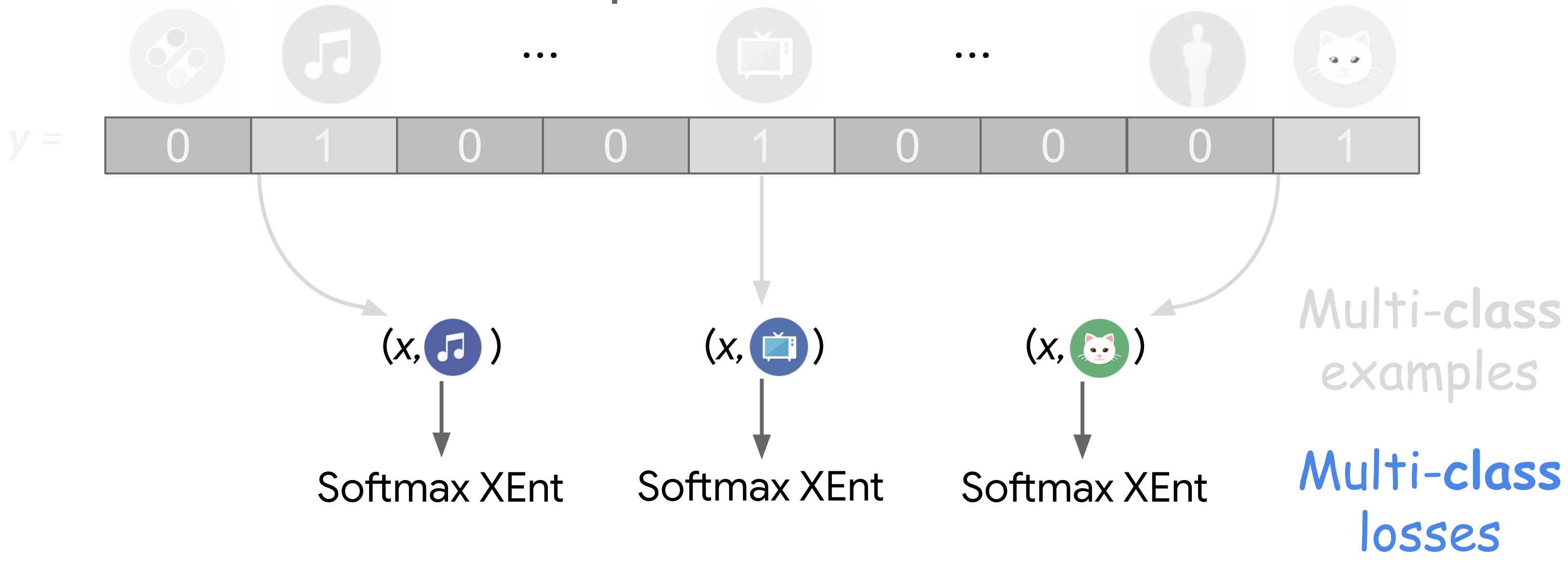
Multilabel algorithm: pick-all-labels

Create a multi-class example for each +'ve label



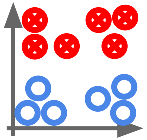
Multilabel algorithm: pick-all-labels

Create a multi-class example for each +'ve label



Multilabel reduction landscape

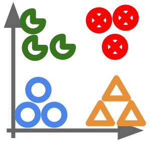
Reductions



Binary

- One-vs-all
- One-vs-all normalised *

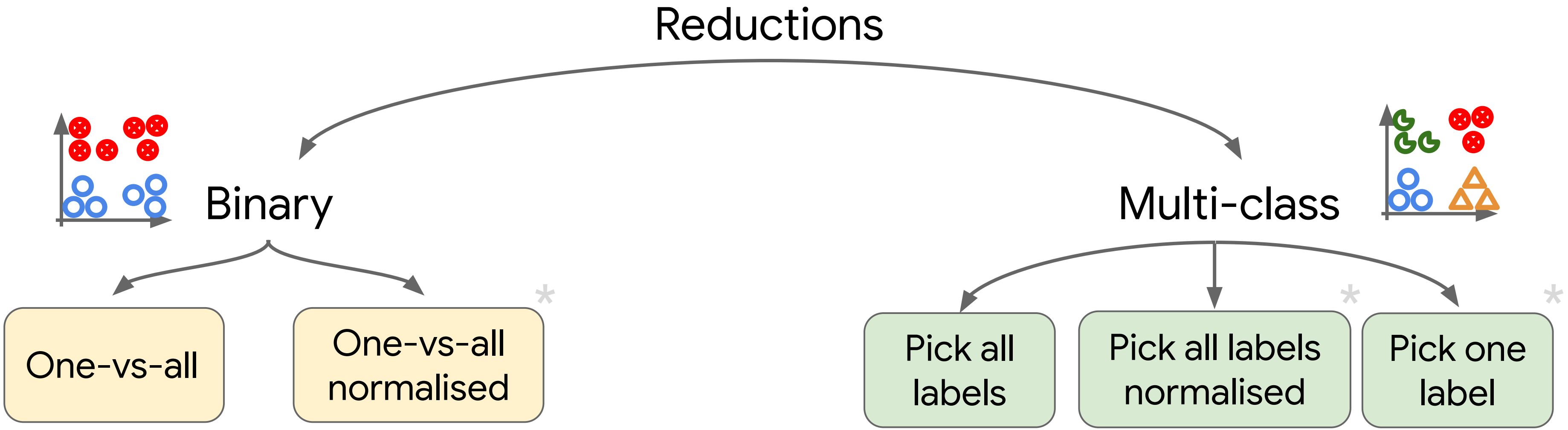
Multi-class



- Pick all labels
- Pick all labels normalised *
- Pick one label *

* See poster for details

Multilabel reduction landscape

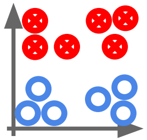


✓ Leverage established losses

* See poster for details

Multilabel reduction landscape

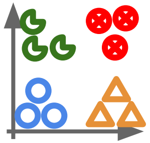
Reductions



Binary

- One-vs-all
- One-vs-all normalised *

Multi-class



- Pick all labels
- Pick all labels normalised *
- Pick one label *

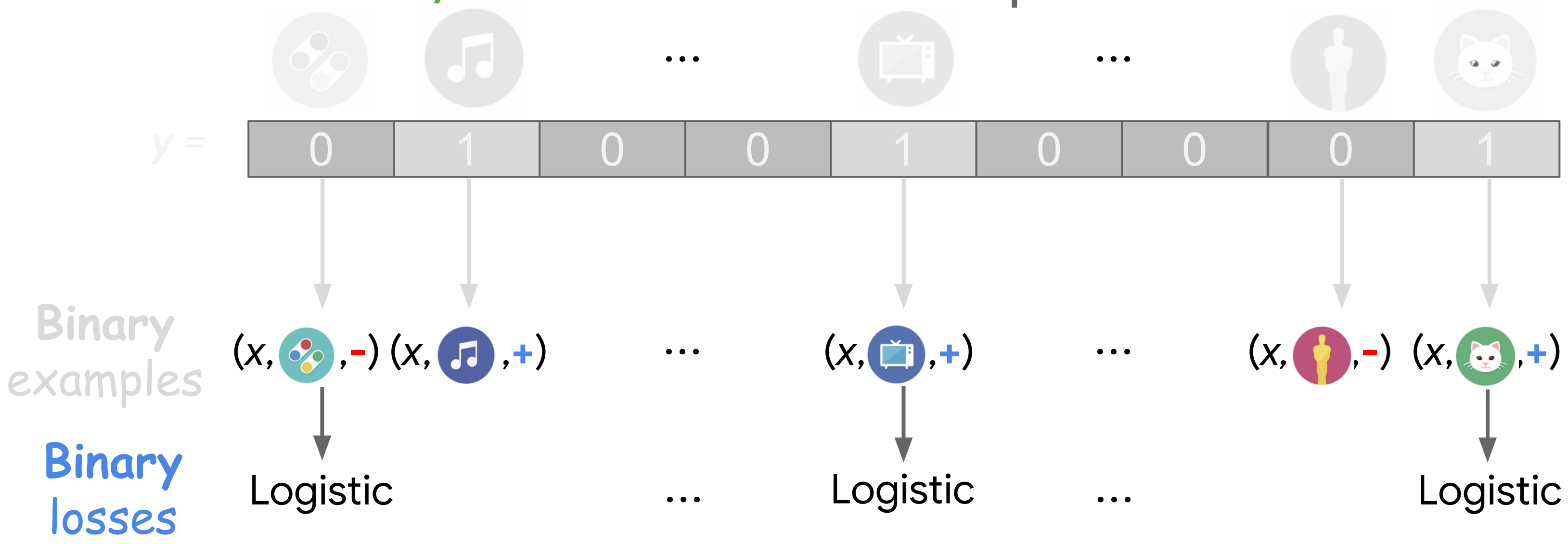
✓ **Leverage established losses**

✗ **What do they optimise?**

* See poster for details

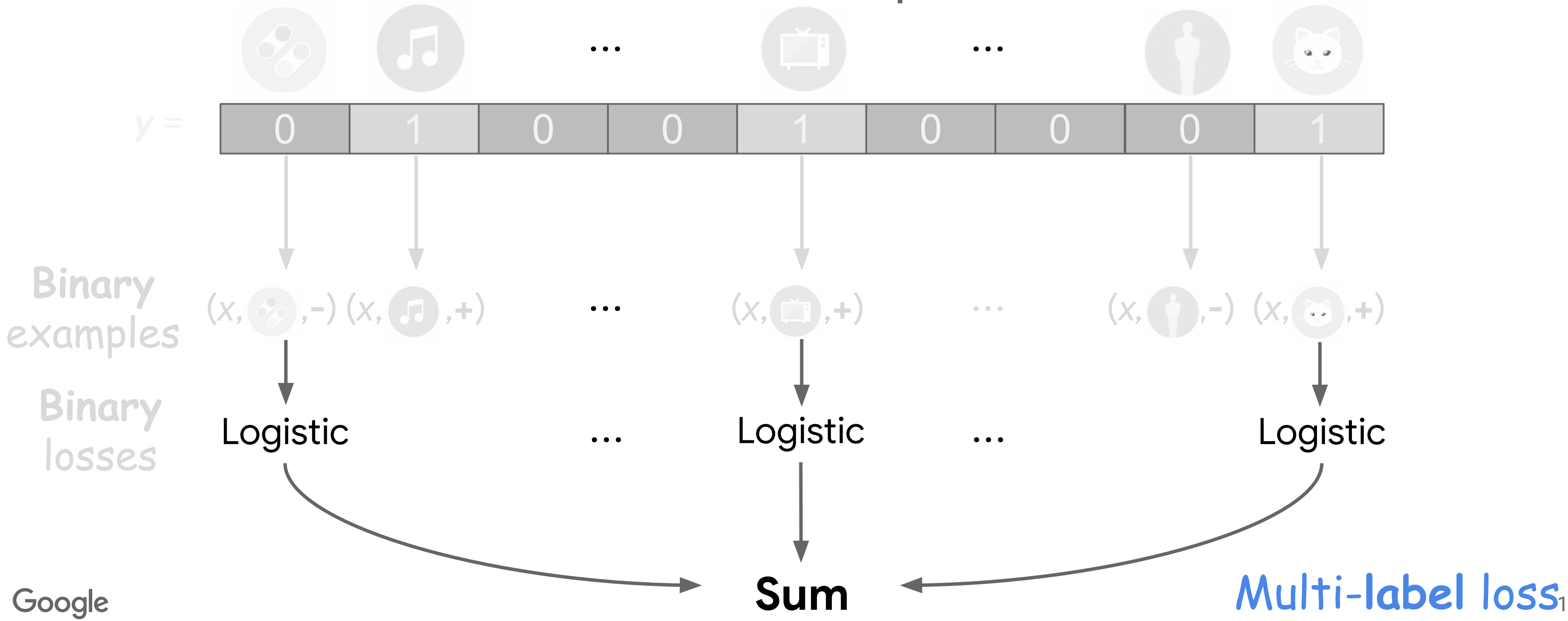
Multilabel loss: one-versus-all

Create a binary-classification example for each label



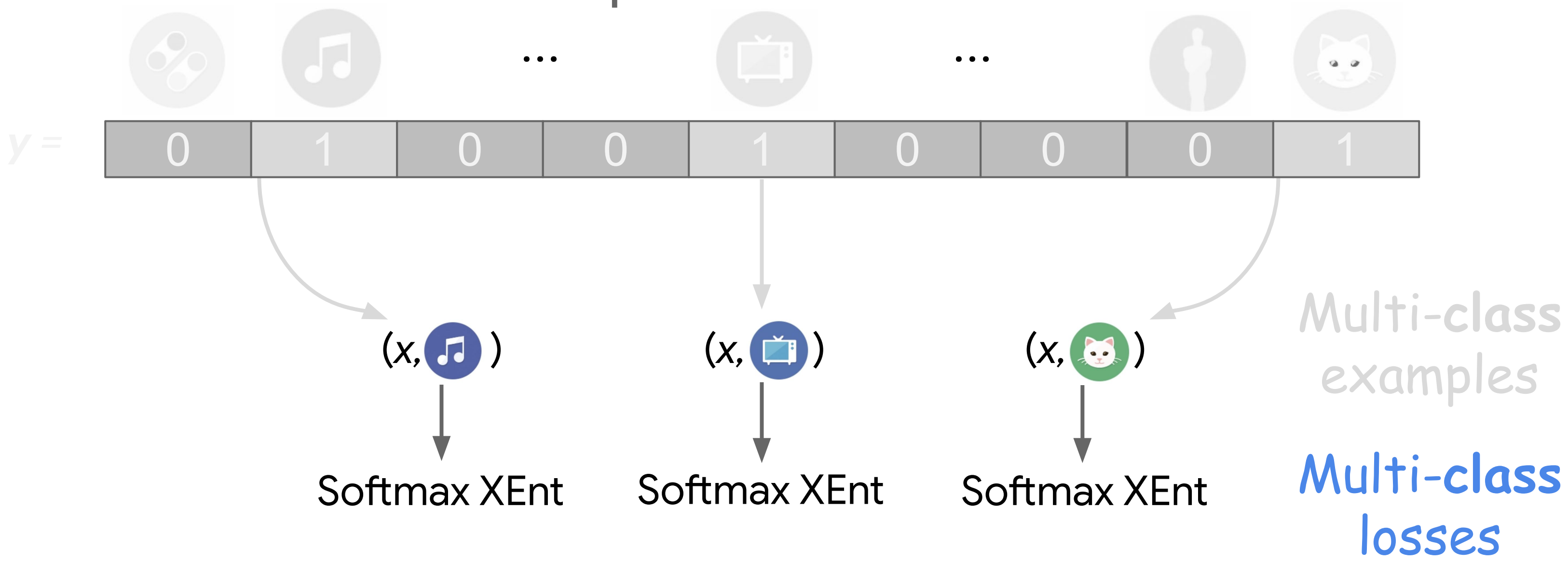
Multilabel loss: one-versus-all

Create a binary-classification example for each label



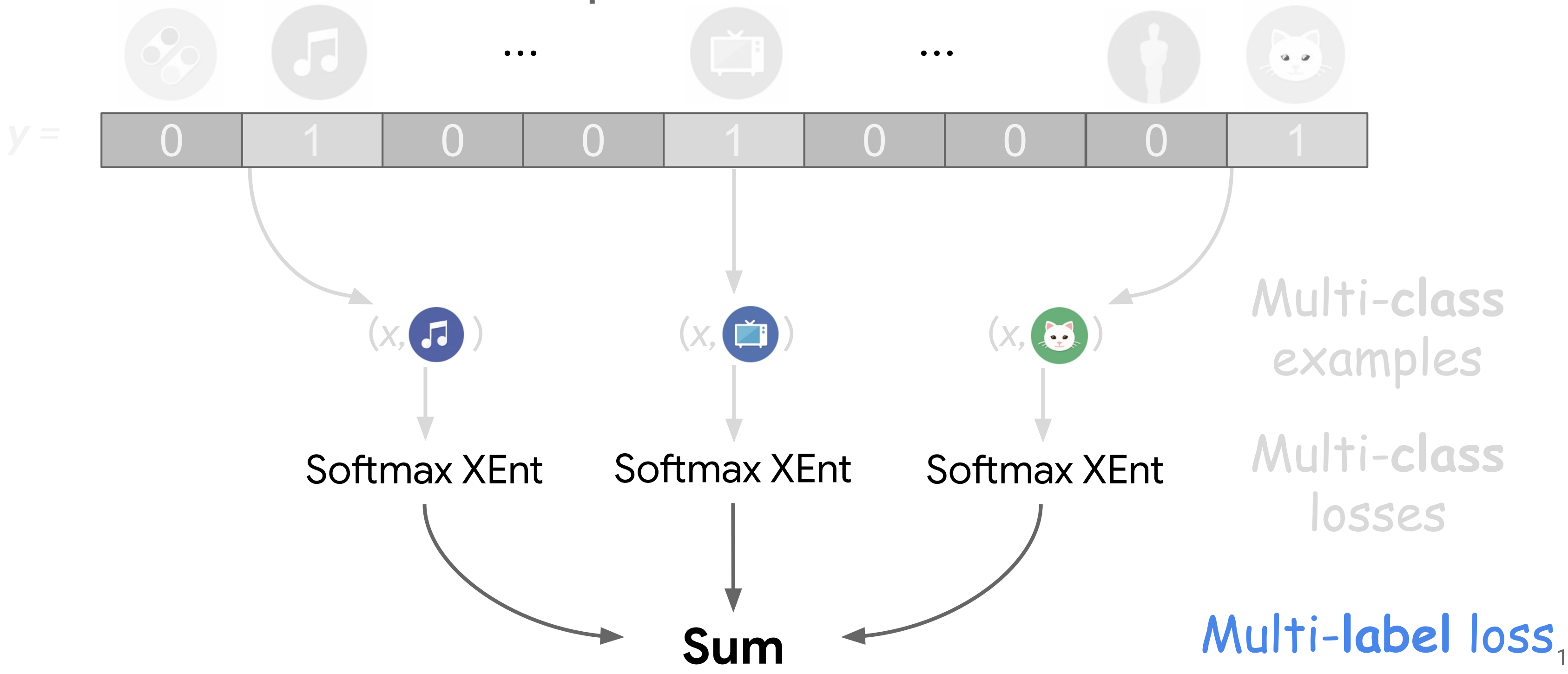
Multilabel loss: pick-all-labels

Create a multi-class example for each +'ve label



Multilabel loss: pick-all-labels

Create a multi-class example for each +'ve label



Bayes-optimal predictions

Bayes-optimal score for *i*th label is:

Reduction	Optimal score
One-versus-all	$\mathbf{P}(y_i = 1 \mid x)$
Pick all labels	$\mathbf{P}(y_i = 1 \mid x) \,/\, N(x)$
All others	$\mathbf{P}(y'_i = 1 \mid x)$

Marginal label relevance

Expected # of +'ve labels

"Normalised" label relevance

Bayes-optimal predictions

Bayes-optimal score for *i*th label is:

Reduction	Optimal score
One-versus-all	$\mathbf{P}(y_i = 1 \mid x)$
Pick all labels	$\mathbf{P}(y_i = 1 \mid x) / N(x)$
All others	$\mathbf{P}(y'_i = 1 \mid x)$

Marginal label relevance

Expected # of +'ve labels

"Normalised" label relevance

"Pick all labels" scores **not coherent** across instances!

Multilabel metrics

Evaluate predictions $f \longrightarrow$ find top-k highest scoring indices

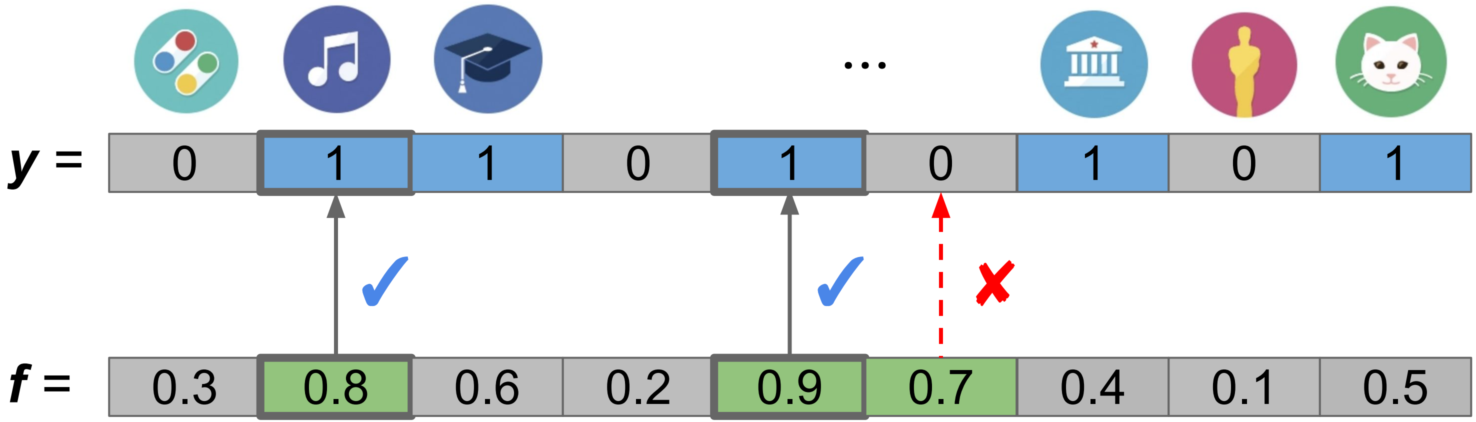
y =	0	1	1	0	1	0	1	0	1

f =	0.3	0.8	0.6	0.2	0.9	0.7	0.4	0.1	0.5

Here, $k = 3$

Multilabel metrics

Evaluate predictions f ⟶ find top-k highest scoring indices

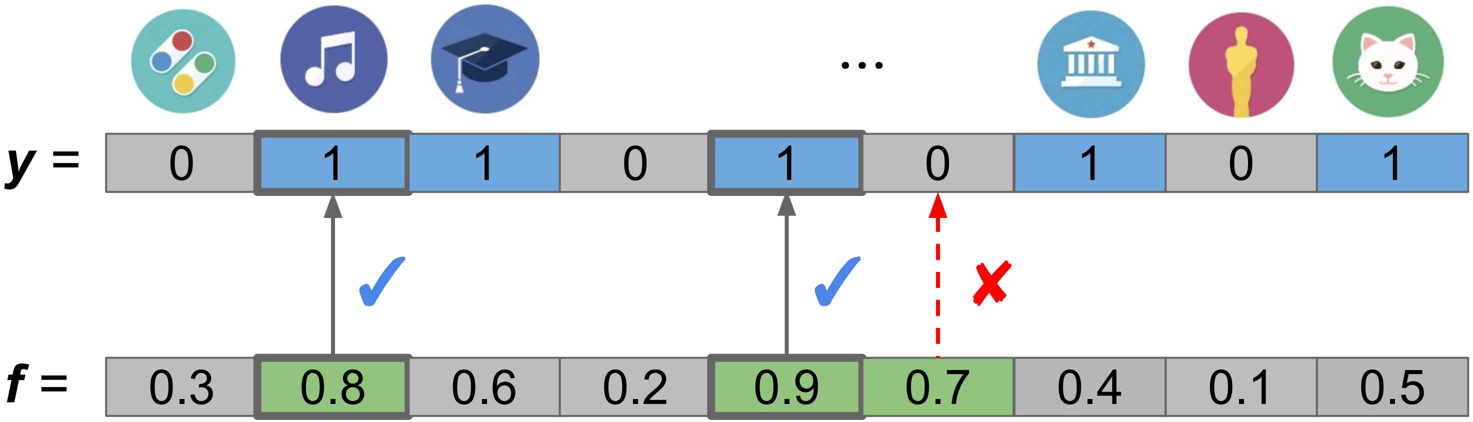


Here, $k = 3$

Precision@k = # positives in top-k / k = ⅔

Multilabel metrics

Evaluate predictions f ⟶ find top-k highest scoring indices



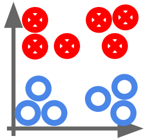
Here, k = 3

Recall@k = # positives in top-k / # positives = ⅖

Google

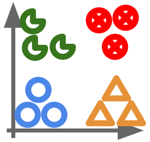
Multilabel reduction landscape

Reductions



Binary

Multi-class



One-vs-all

One-vs-all normalised

Pick all labels

Pick all labels normalised

Pick one label

✓ Leverage established losses

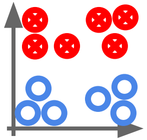
✘ What do they optimise?

Multilabel reduction landscape

Consistent for **precision@*k***

Consistent for **recall@*k***

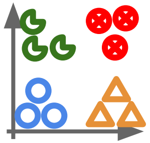
Reductions



Binary

- One-vs-all
- One-vs-all normalised

Multi-class



- Pick all labels
- Pick all labels normalised
- Pick one label

✓ **Leverage established losses**

✓ **Consistent for precision <u>or</u> recall**

Drop by East Exhibition Hall B + C #14

Unified analysis of distinct multilabel reductions

Consistency for precision@k or recall@k -- but not both!