Multilabel reductions

What is my loss optimising?

Aditya Krishna Menon

Ankit Singh Rawat

Sashank Reddi

Sanjiv Kumar

Google Research

Multilabel learning

Given an instance (e.g., video), predict a binary label vector (e.g., categories):

1 1 0

0 1 1

Multilabel metrics

Multi-label observations: $(x, \boldsymbol{y})$ where $\boldsymbol{y} \in \{0, 1\}^L$

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

Canonical measures of performance: for integer k,

Precision@k = $\mathbb{E}_{(x, \boldsymbol{y})}[\, | \text{Top}_k(f(x)) \cap \boldsymbol{y} | \,/\, k\,]$

Recall@k = $\mathbb{E}_{(x, \boldsymbol{y})}[\, | \text{Top}_k(f(x)) \cap \boldsymbol{y} | \,/\, |\boldsymbol{y}|\,]$

Differ in denominator only

Number of relevant labels in top-k predictions

Multilabel losses

Need to specify suitable loss function $\ell(y, f(x))$ while training

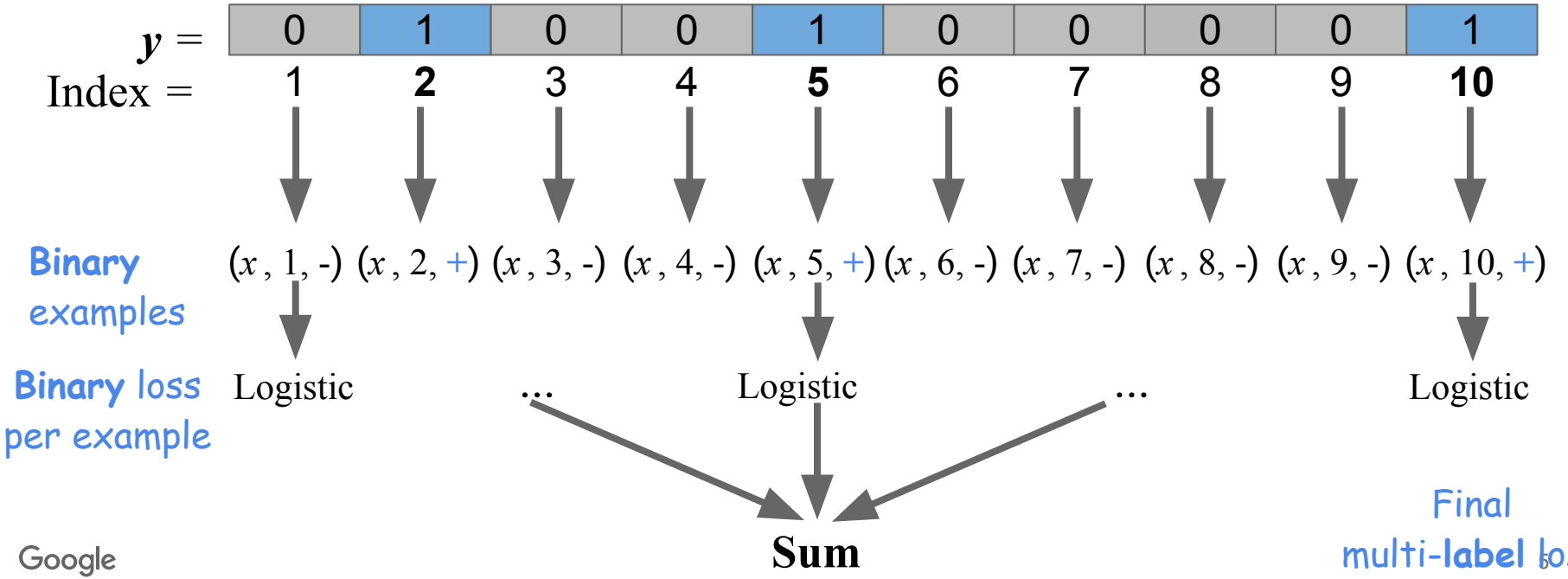
Ideally, want to leverage binary and multi-class losses

- Hinge, softmax cross-entropy, ...

Can achieve this using a reduction!

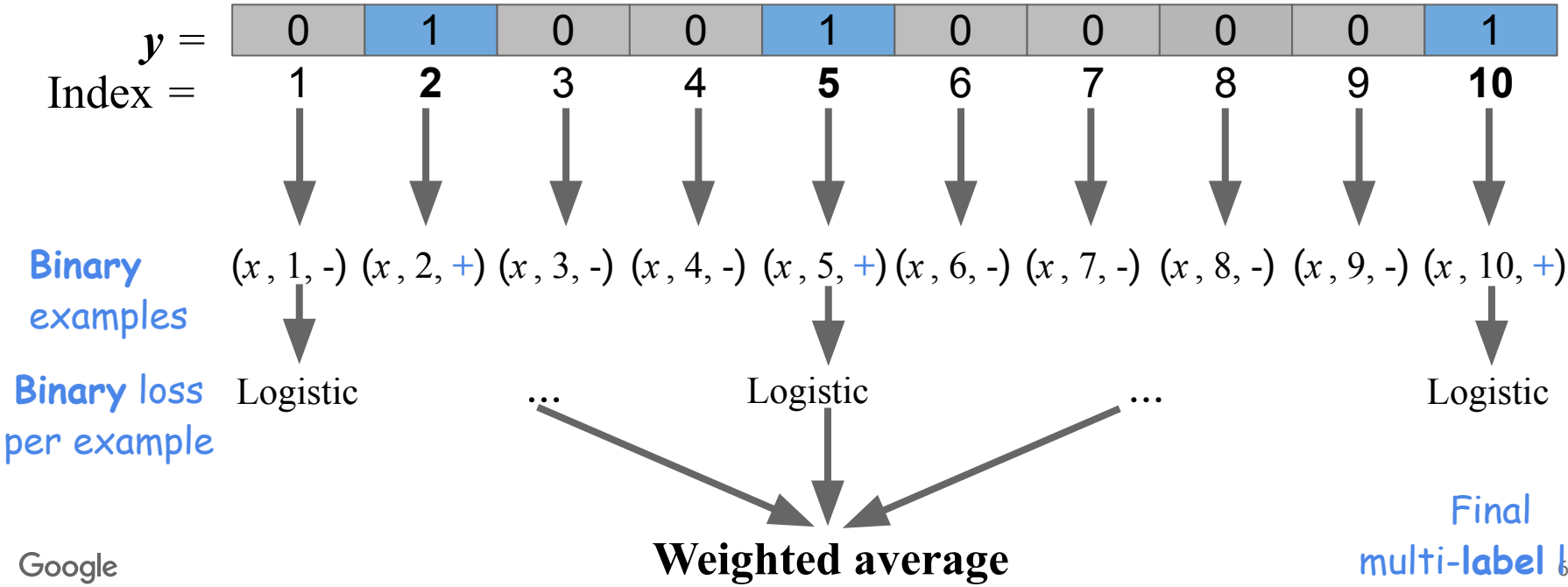
Reduction: one-versus-all

Given (x, y), create a binary-classification example for each label



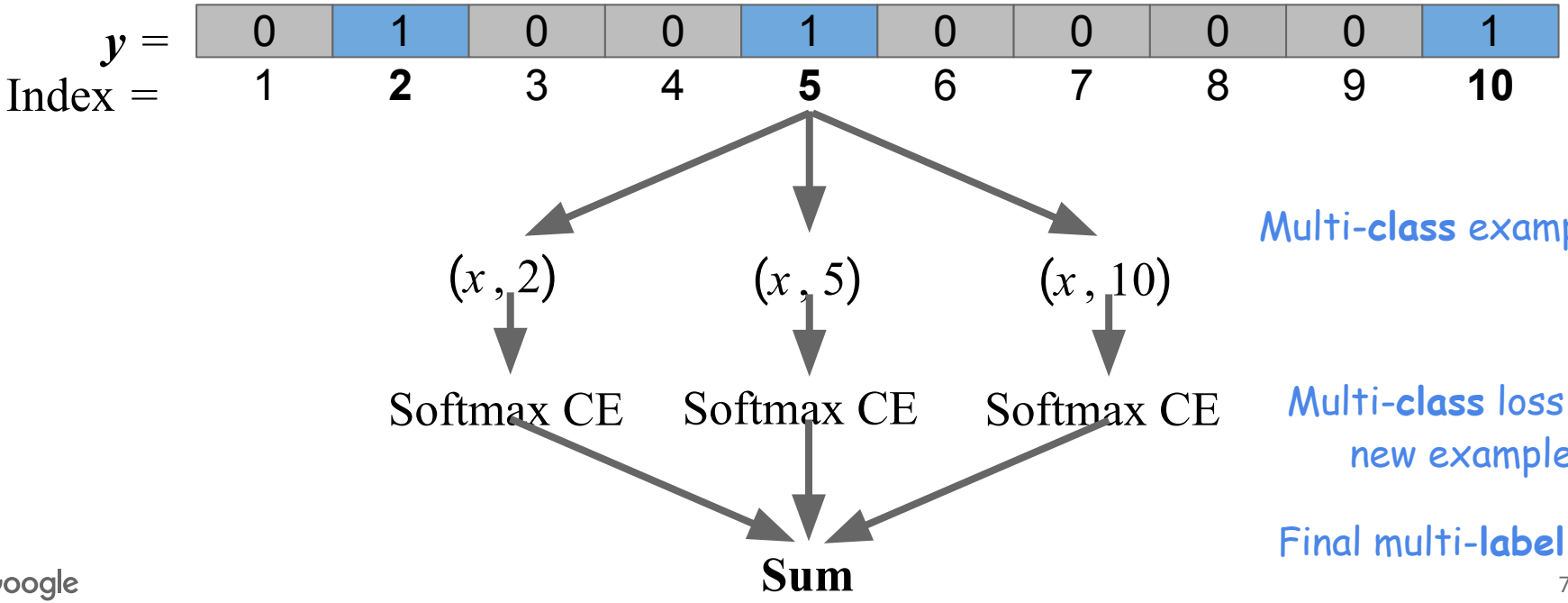
Reduction: one-versus-all normalised

Given (x, y), create a binary-classification example for each label, and weight inversely by total # of positive labels



Reduction: pick-all-labels

Given (x, y), create a multi-class example for each positive label



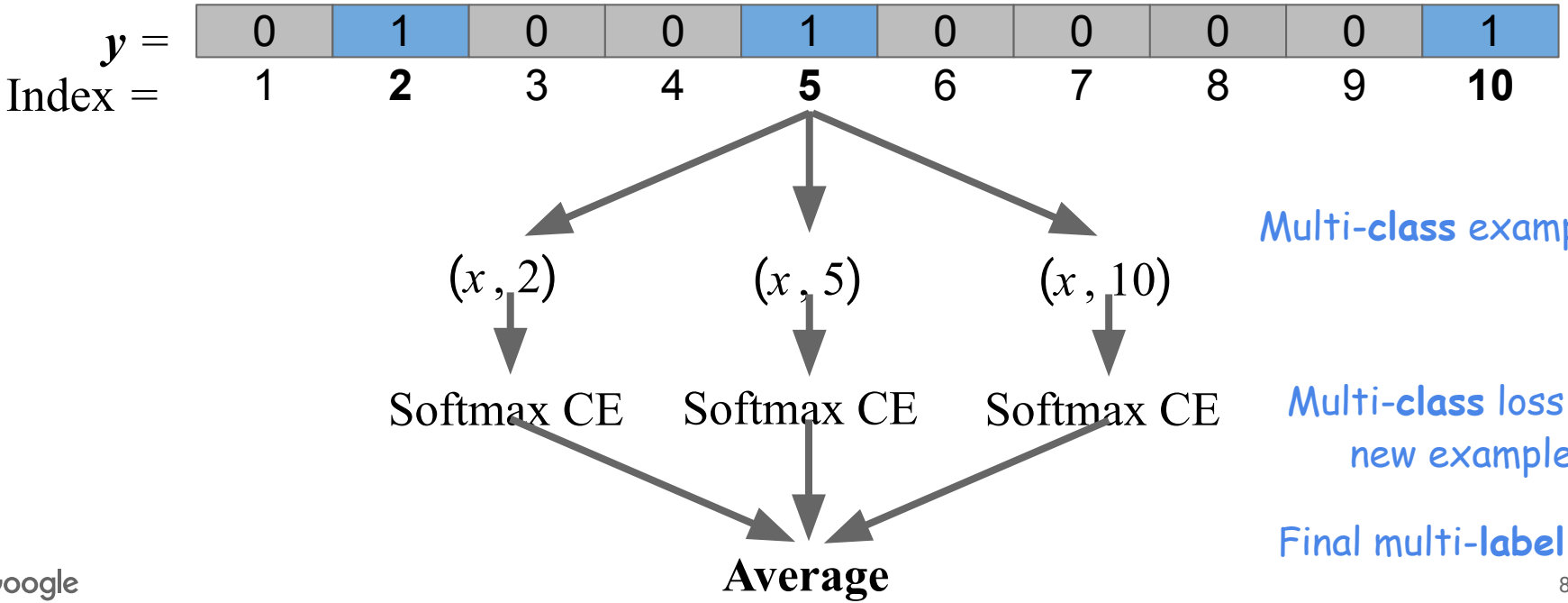
Multi-**class** examples

Multi-**class** loss per new example

Final multi-**label** loss

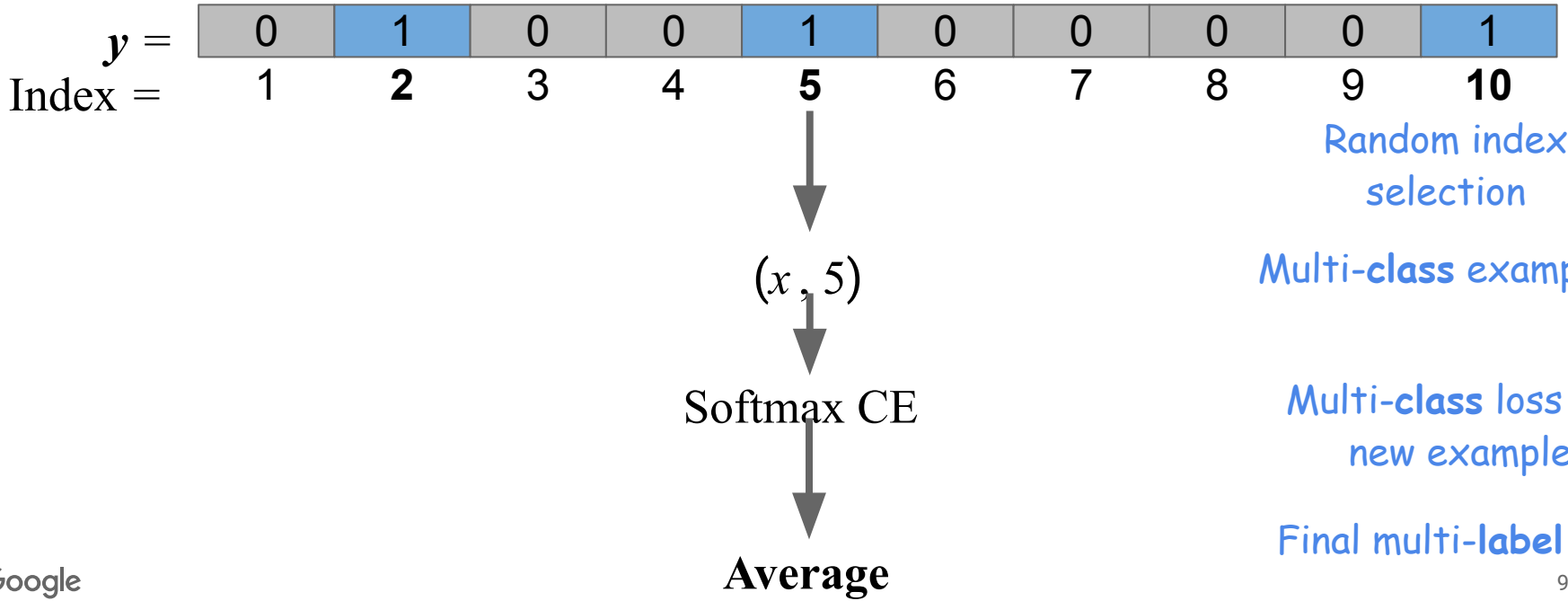
Reduction: pick-all-labels normalised

Given (x, y), create a multi-class example for each positive label, and weight inversely by total # of positive labels

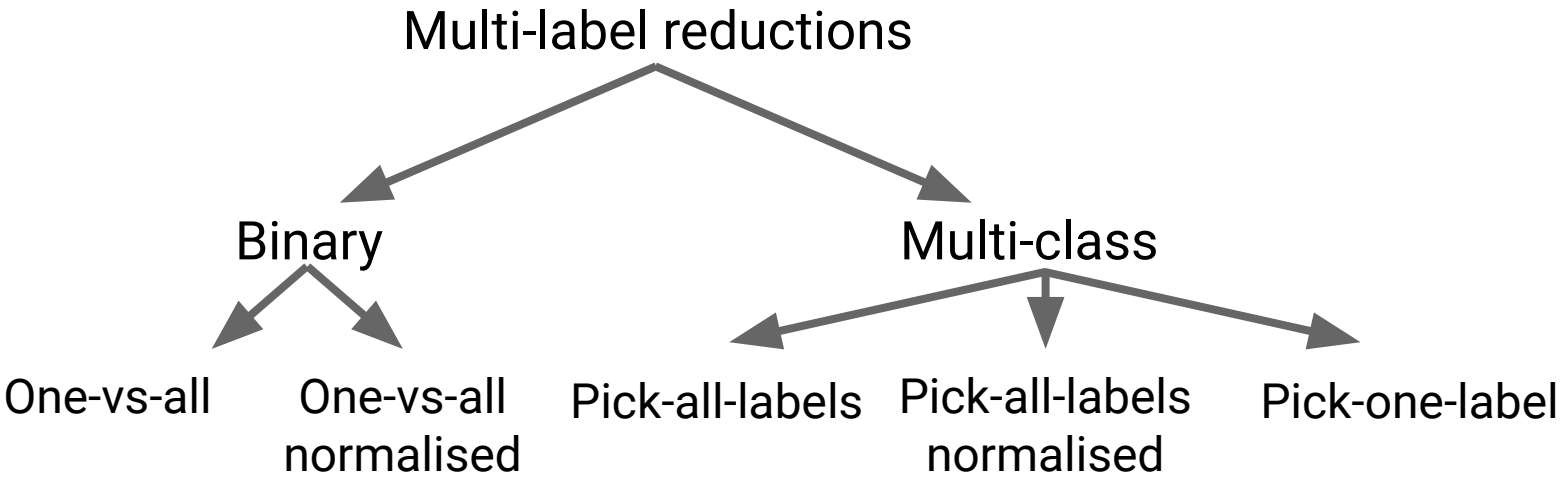


Reduction: pick-one-label

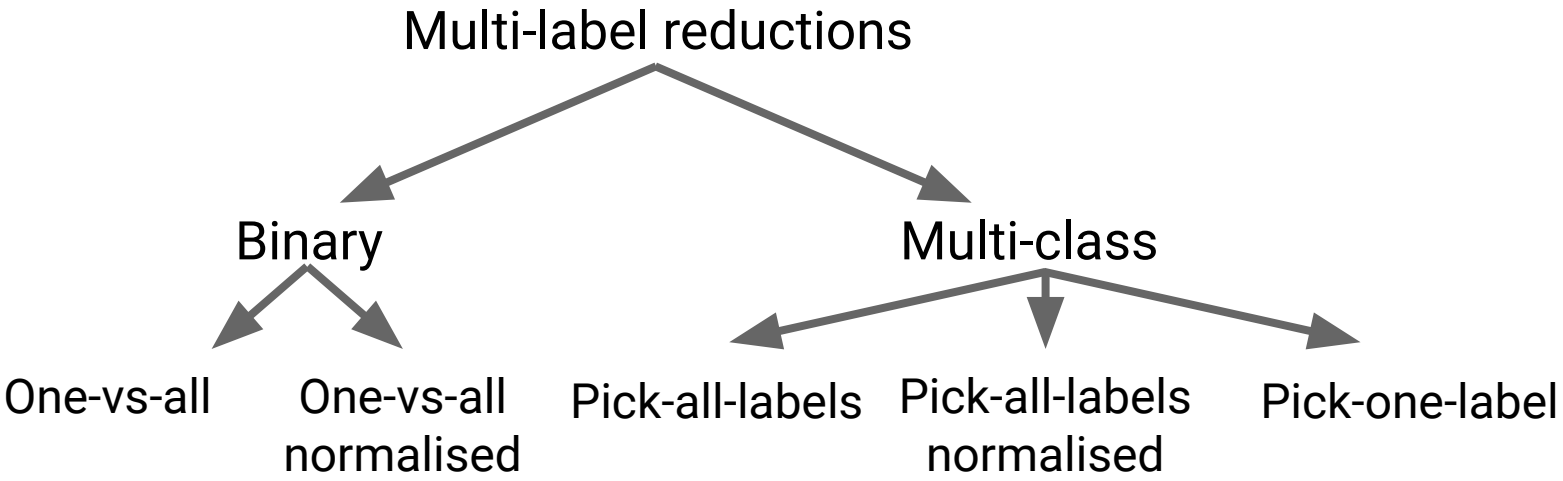
Given (x, y), create a single multi-class example by randomly drawing a single positive label



Multilabel reductions: overview

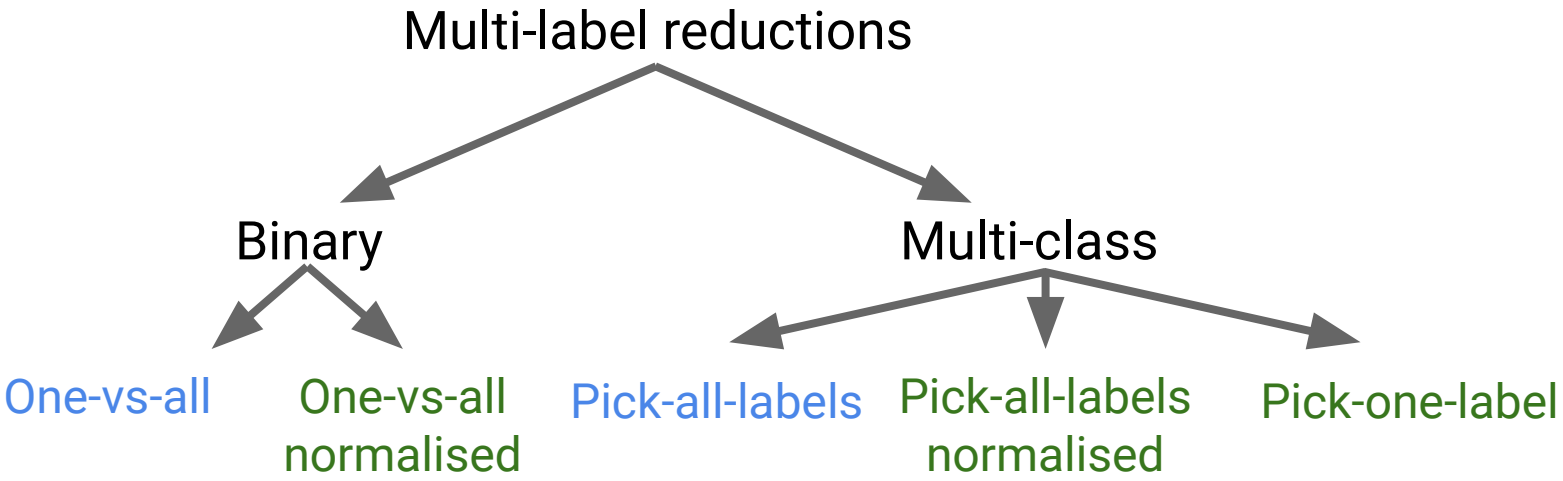


Multilabel reductions: overview



Intuitive, but what do they optimise?

Multilabel reductions: what are we optimising?



Each reduction is consistent for either precision@k or recall@k -- but **not both**!

- choose reduction based on one's ultimate goal!

Thanks!

Paper ID #5615