

Class-Distribution Regularized Consensus Maximization for Alleviating Overfitting in Model Combination

Sihong Xie, Jing Gao, Wei Fan, Deepak Turaga and Philip S. Yu

University of Illinois at Chicago
University at Buffalo
Huawei Noah Lab
IBM T.J Watson Research

UIC

UNIVERSITY OF ILLINOIS
AT CHICAGO

UB

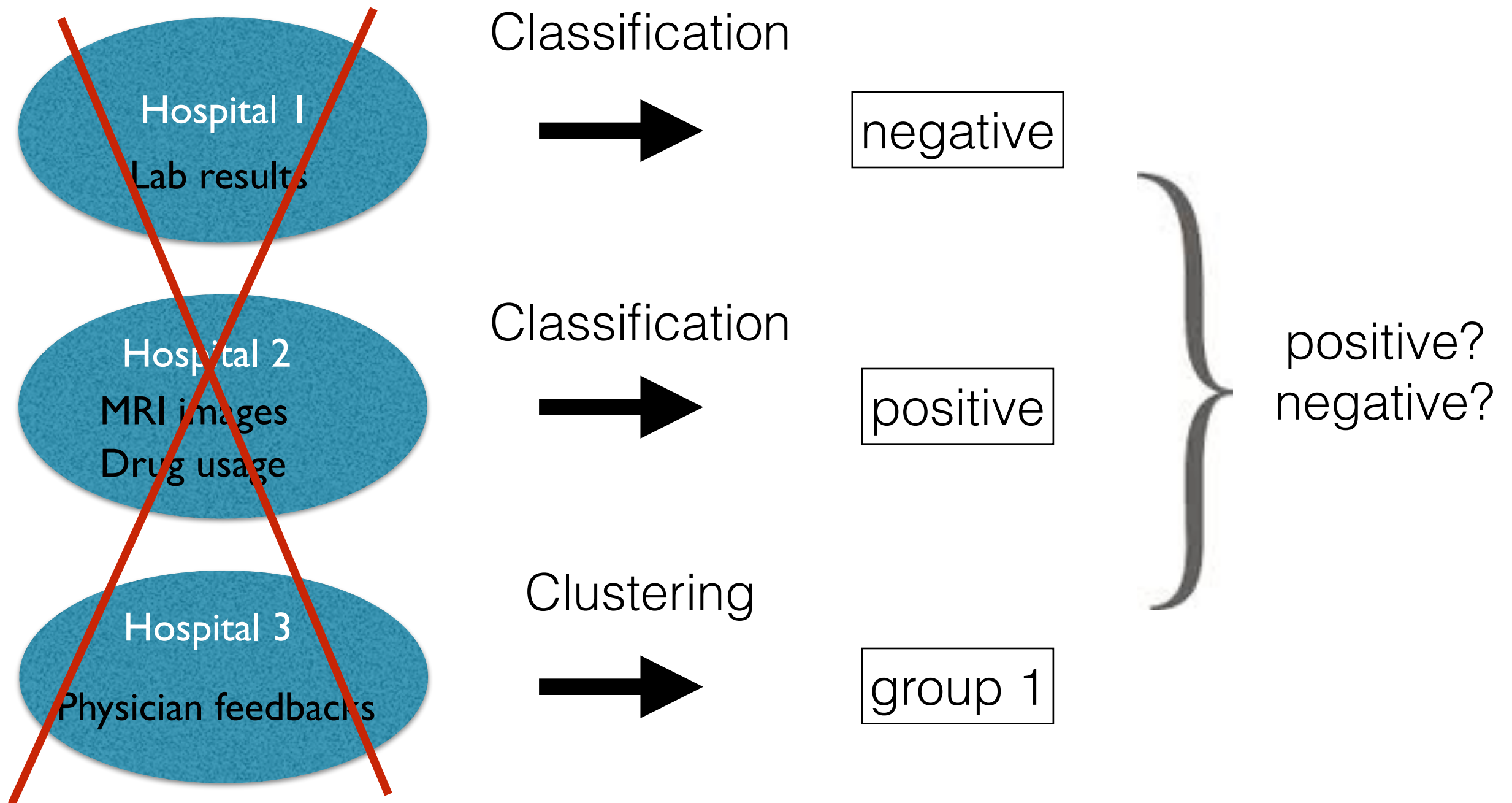
University at Buffalo
The State University of New York



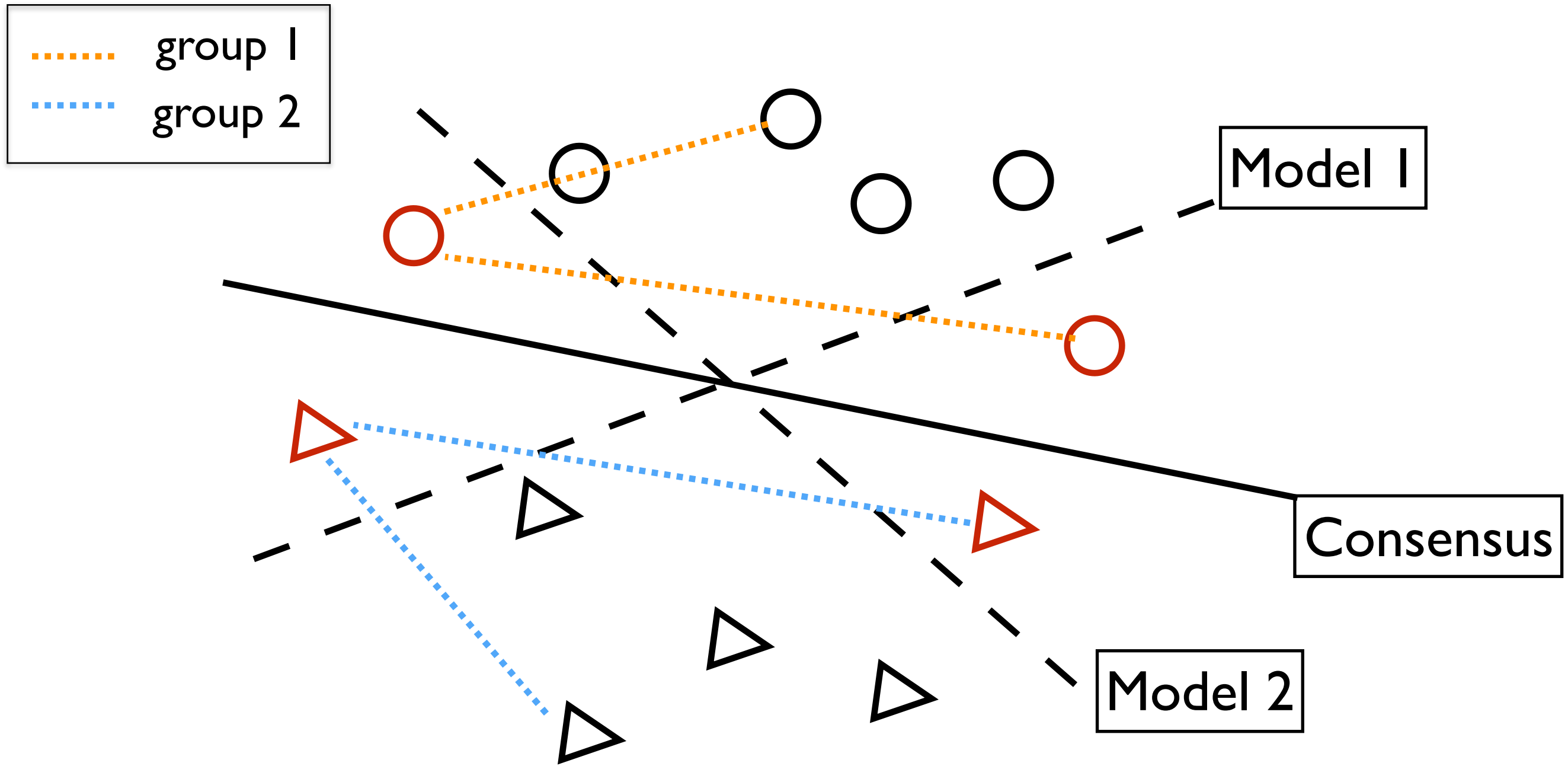
HUAWEI

IBM
Research

Prediction Combination: a motivating example



Prediction combination

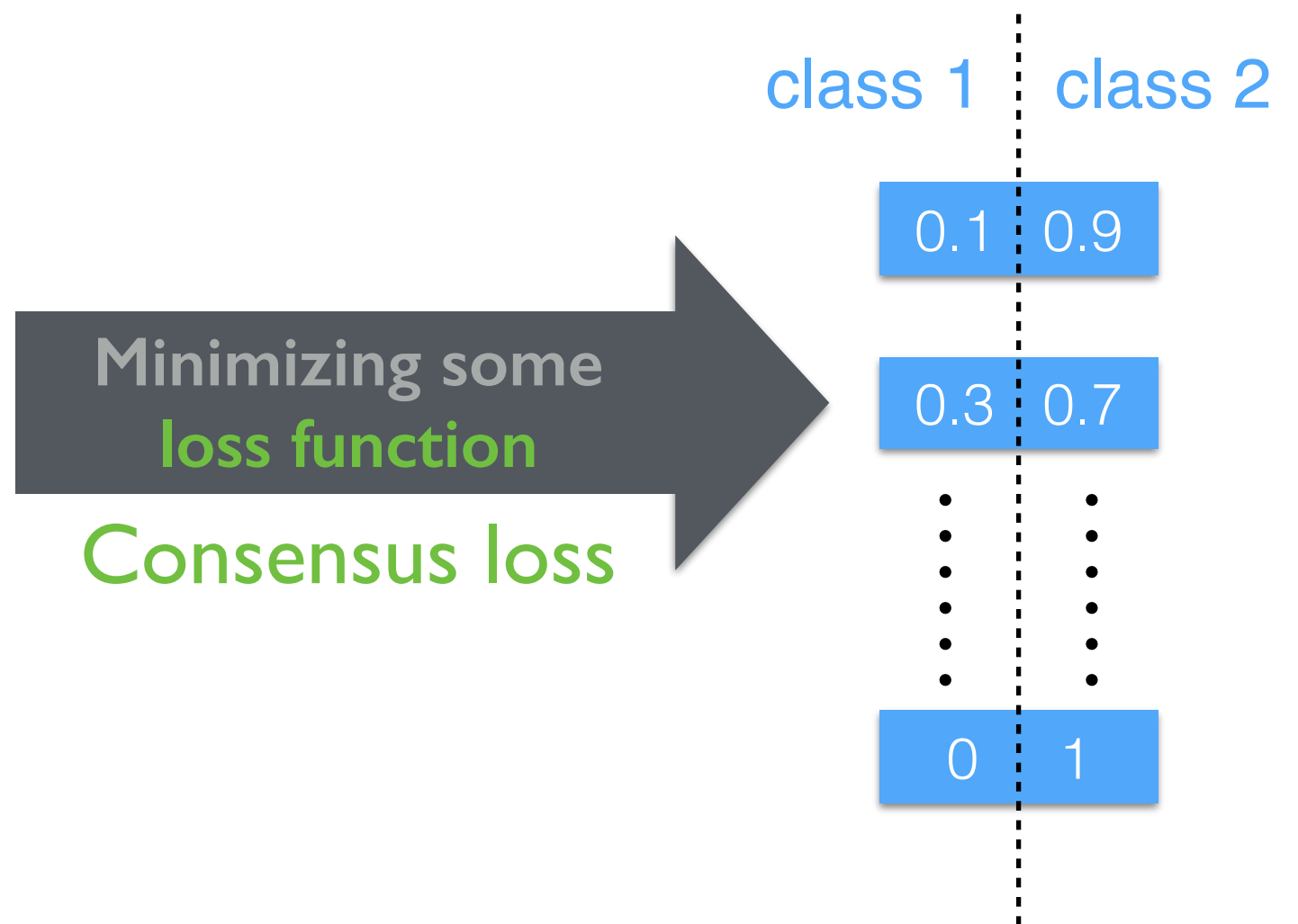


Prediction combination: a high level view

Multiple models' predictions

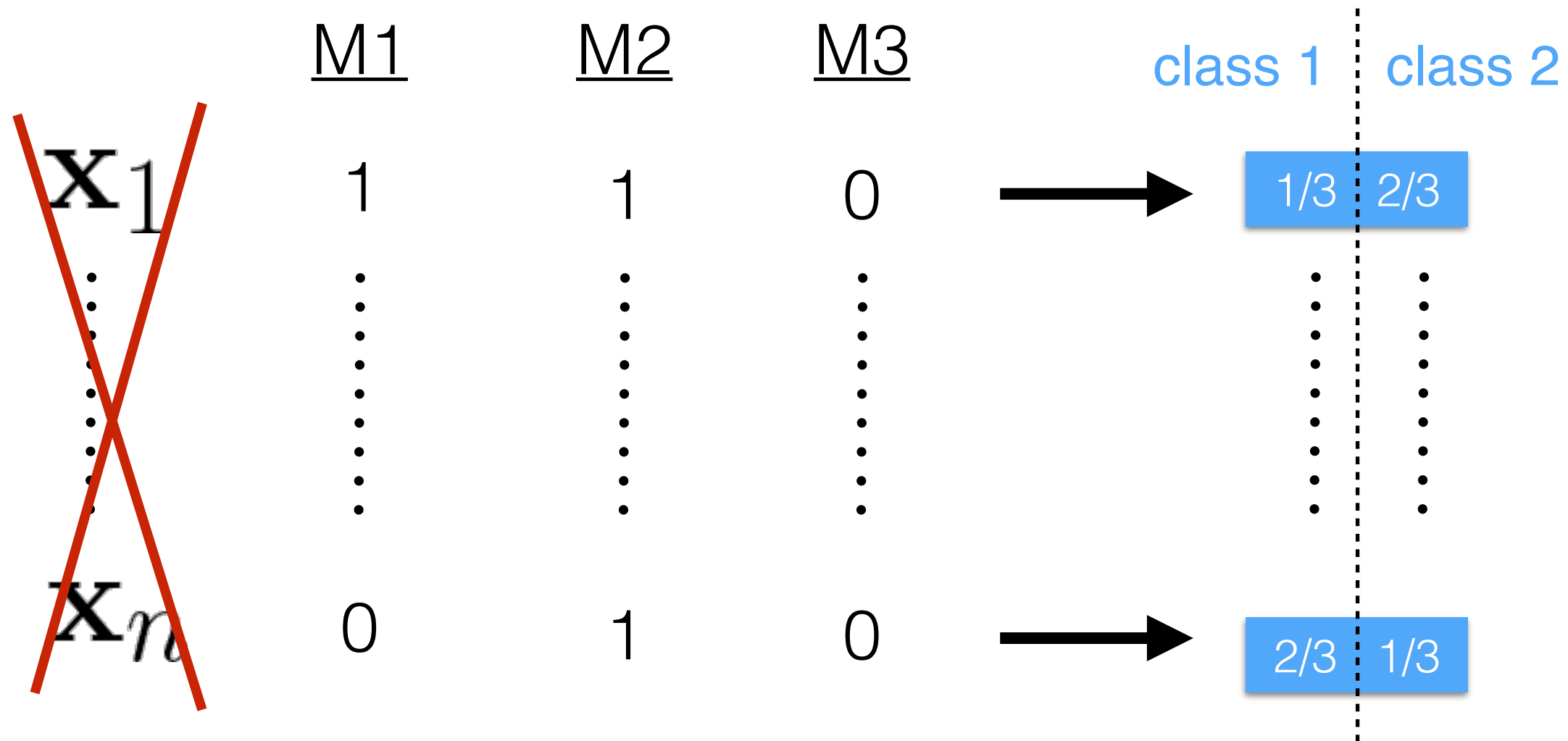
	<u>M1</u>	<u>M2</u>	<u>M3</u>
X_1	1	1	0
\vdots	\vdots	\vdots	\vdots
X_n	0	1	0

Combined predictions



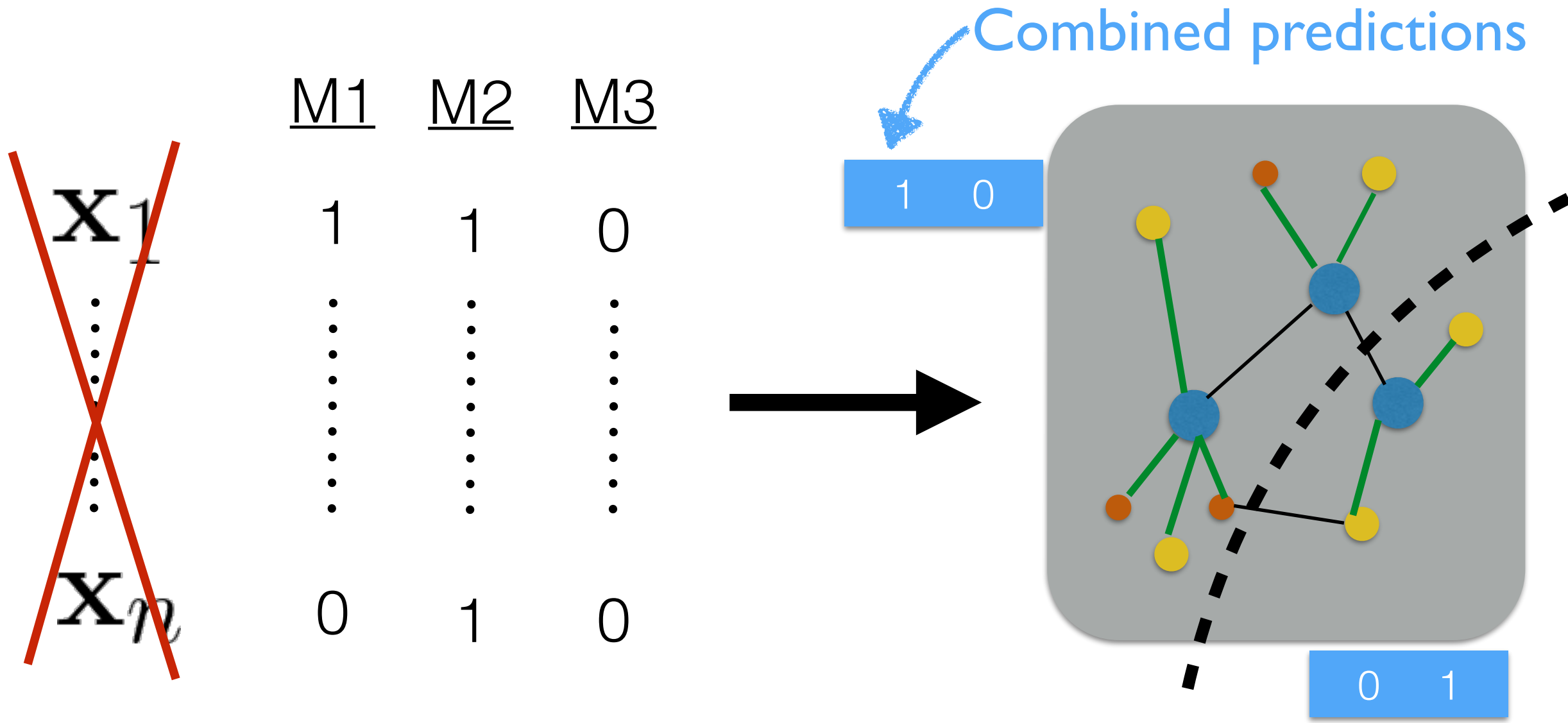
Traditional approaches: Majority voting

Combined prediction



Consensus loss: mean squared error loss

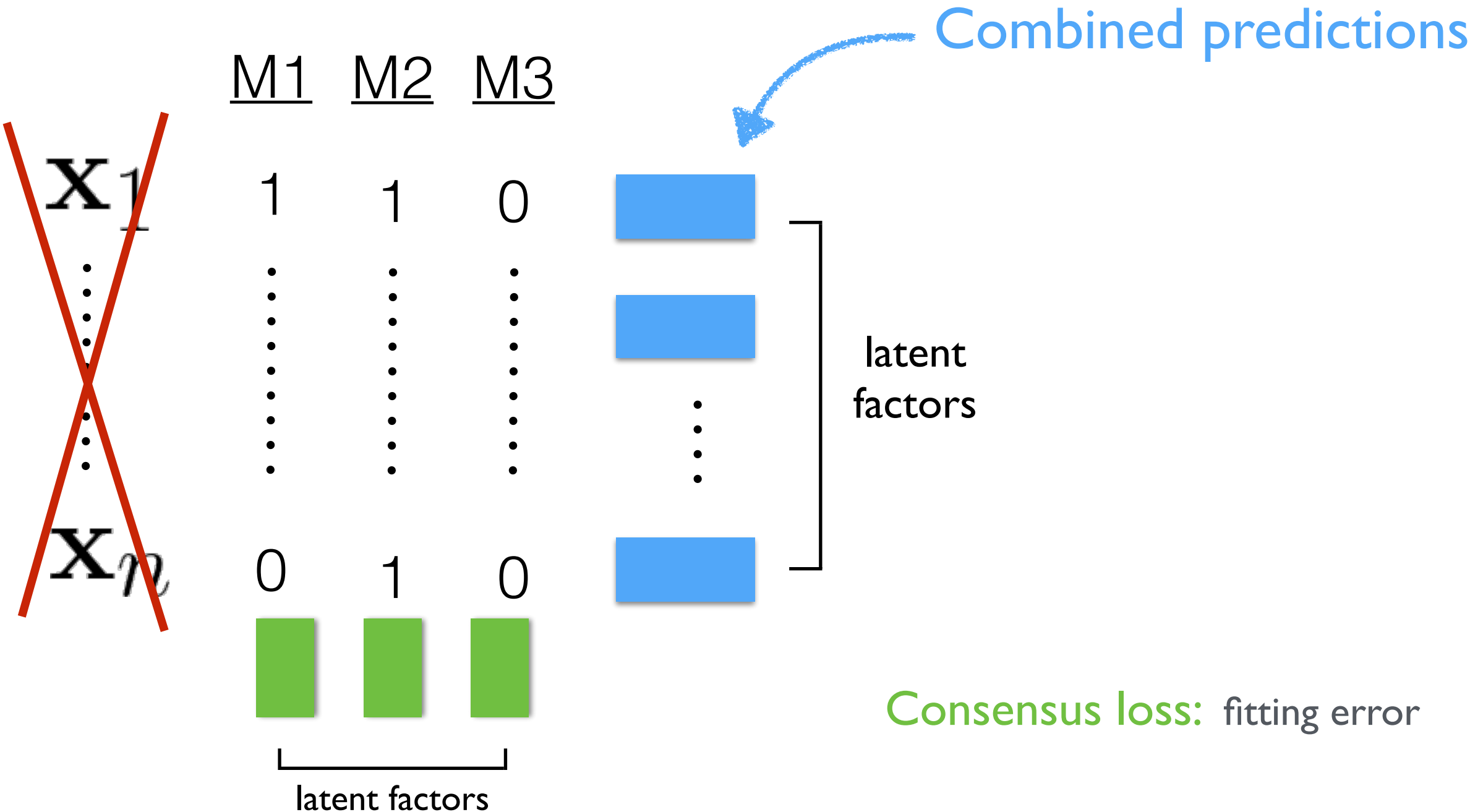
Traditional approaches: Graph-based model



Consensus loss: volume of cuts of the graphs

Representative methods: HGPA, HBGF, MCLA, Consensus maximization

Traditional approaches: factorization

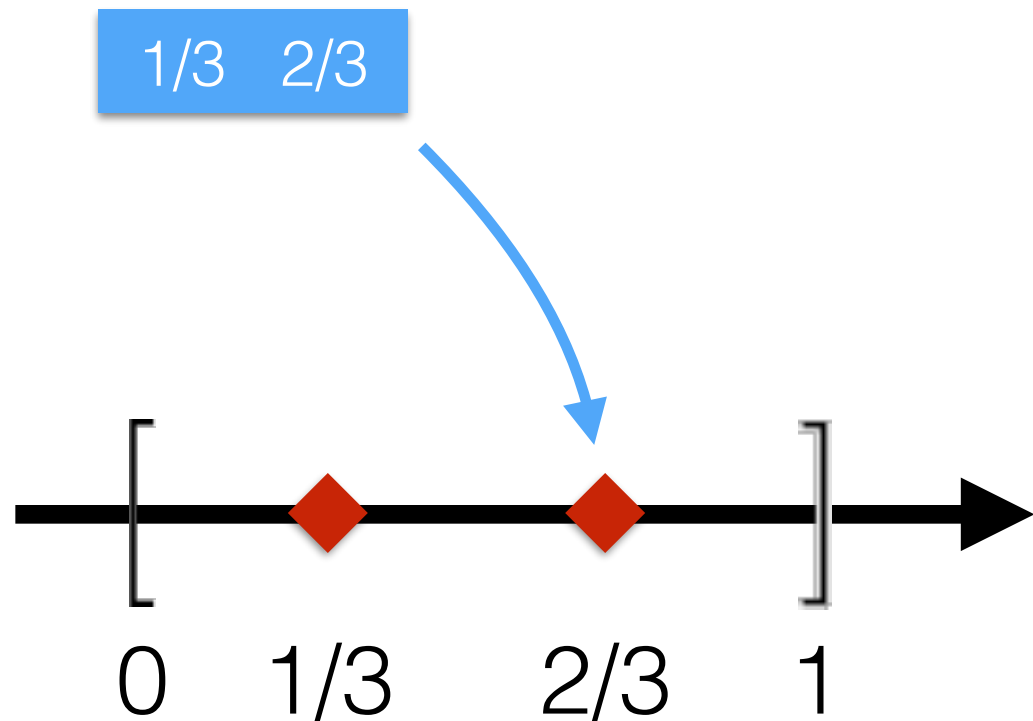


Representative methods: Non-negative MF, LDA

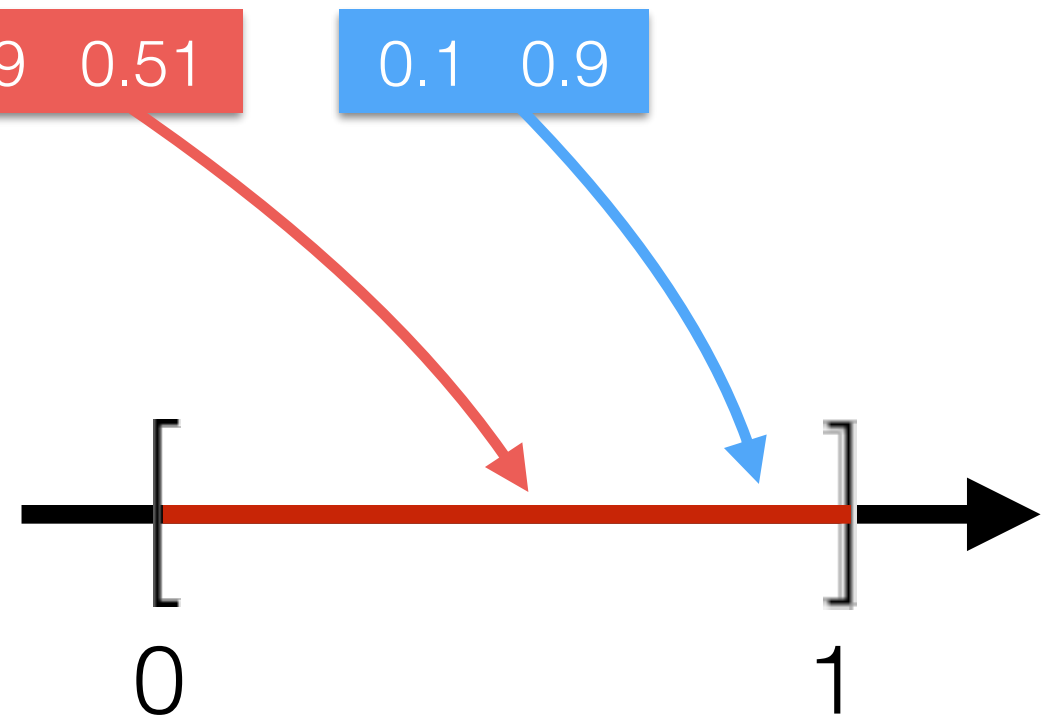
Shortcomings of traditional approaches

Majority voting

consensus maximization

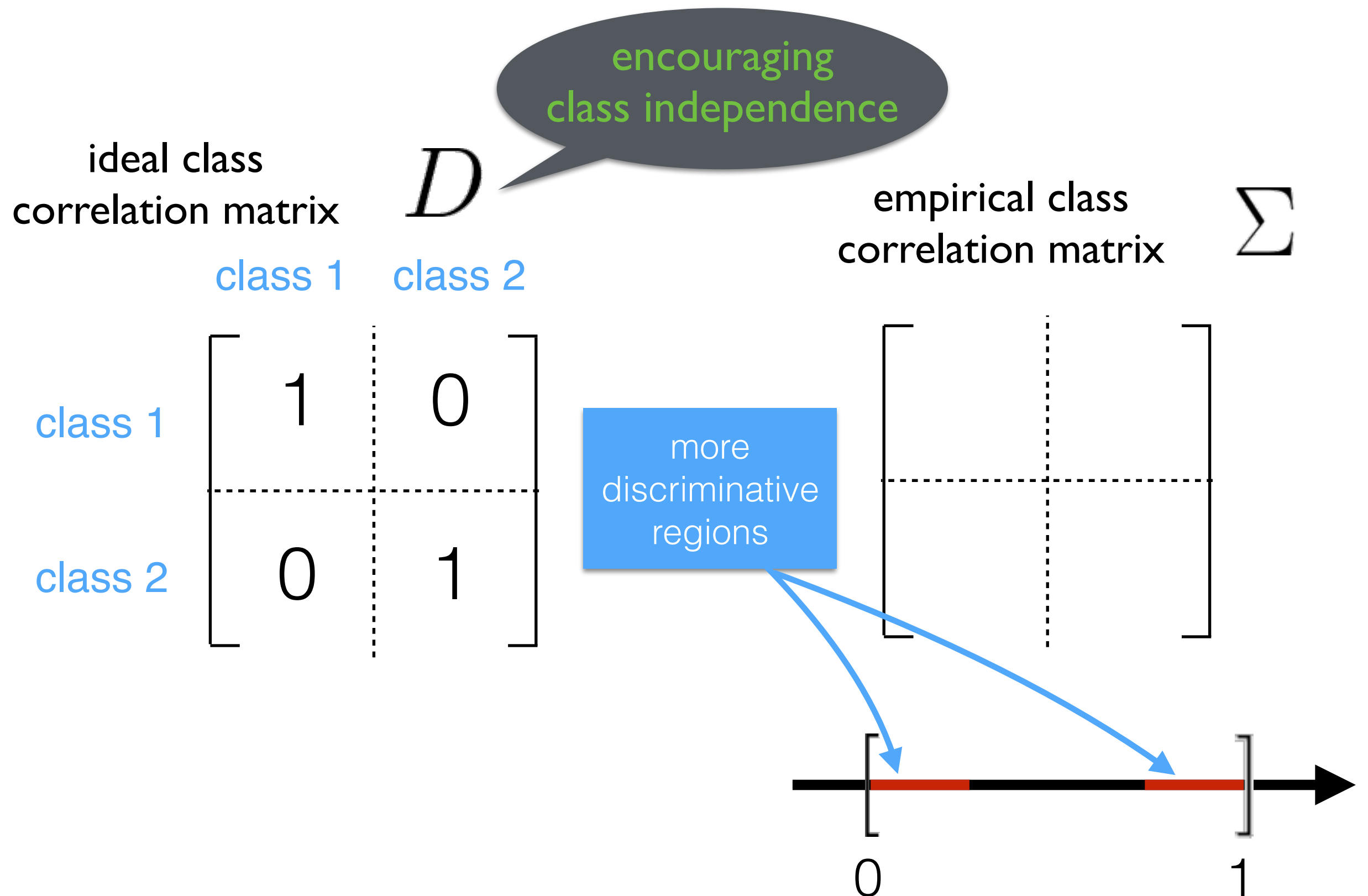


Only allows a subset of rationals
Too restricted!



The whole interval
May lead to **overfitting!**

Regularized Solutions



Regularized Solutions

Framework:

minimize: consensus loss + regularization term

- *mean squared error*
- *graph cut*
- *negative log-likelihood*

- *Euclidean Distance(D, Σ)*
- *KL divergence(D, Σ)*

An instantiation:

minimize: graph cut loss + Euclidean distance

Optimization: alternative block-wise gradient descent

Theoretical guarantee: achieving a smaller generalization upper bound

Experiments

Datasets: 11 multi-class text classification tasks

Base models: 2 classifiers + 2 clustering models

Prediction combination baselines:

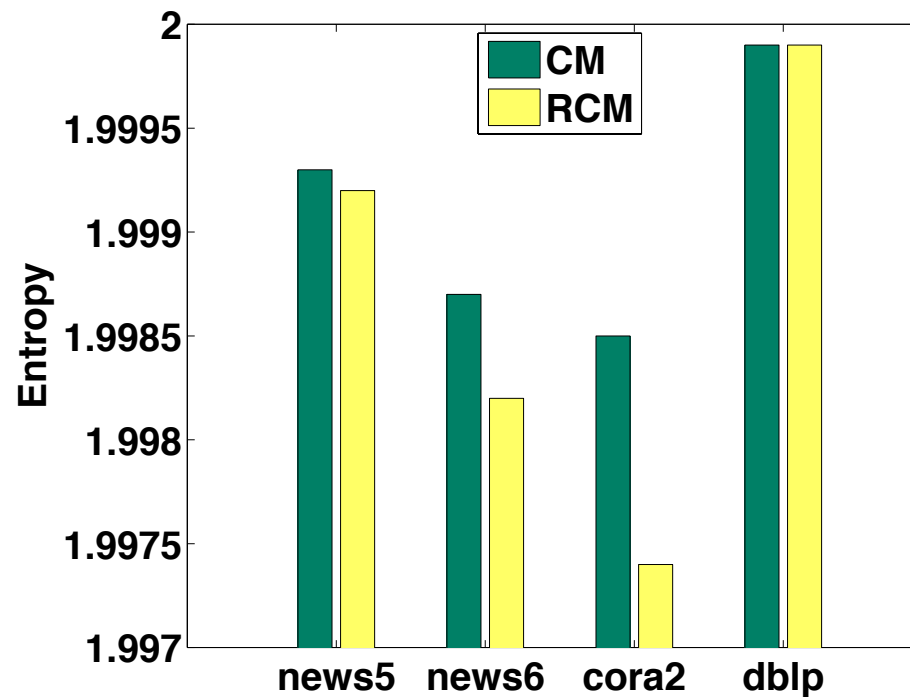
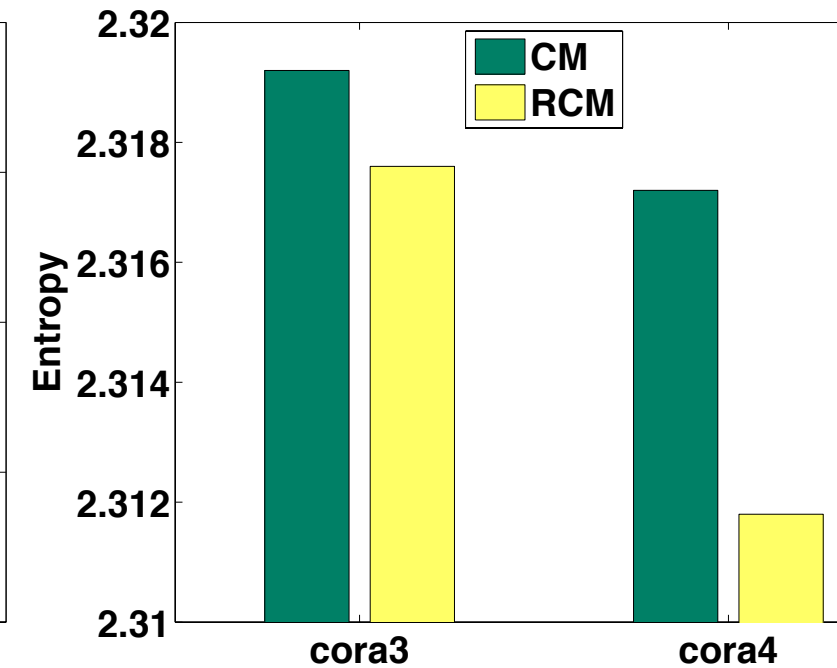
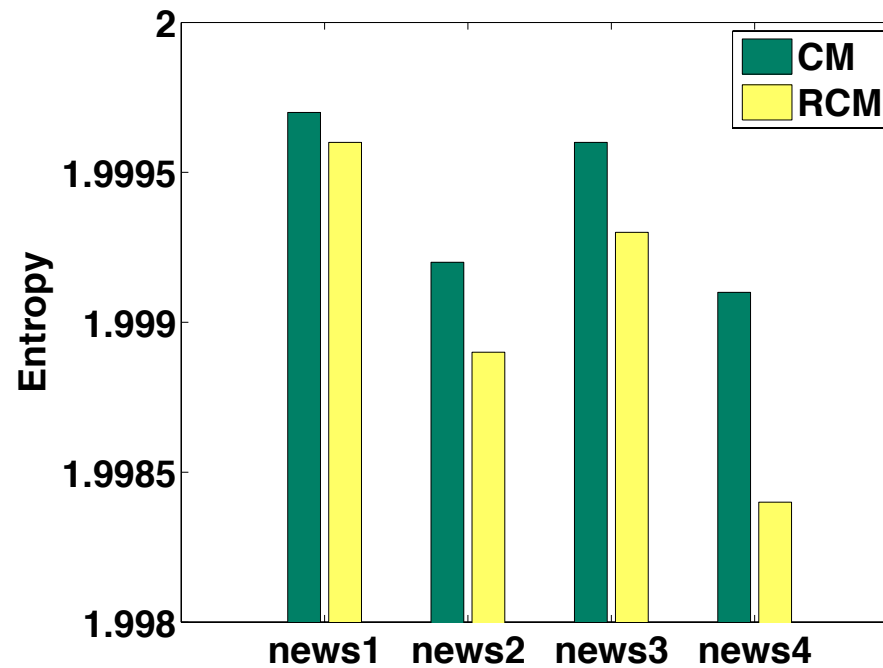
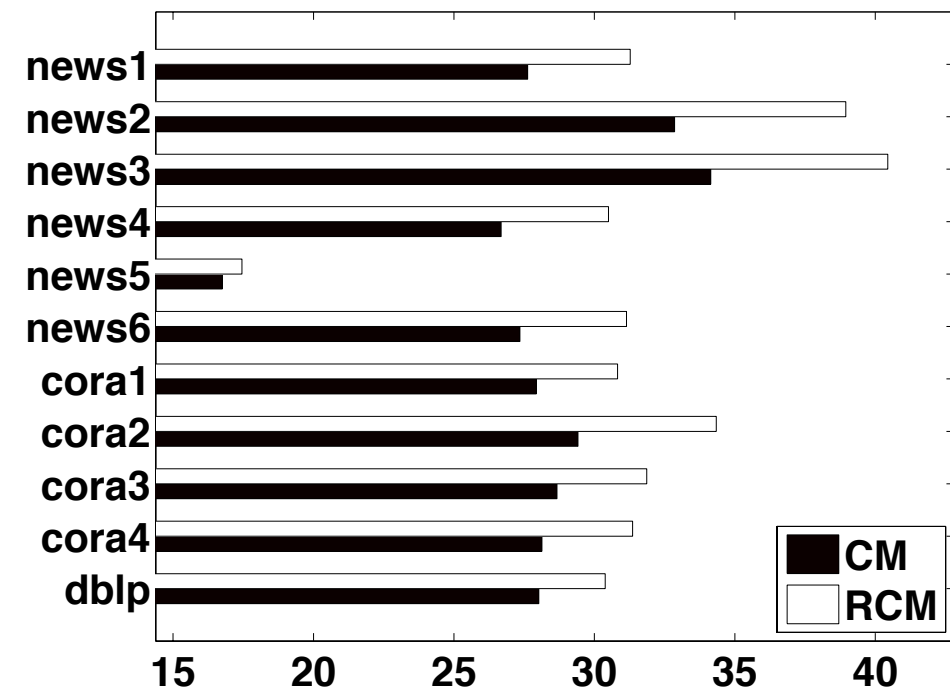
- *majority voting*
- *graph-cutting methods: consensus maximization, HBGF, MCLA*
- *factorization-based: Bayesian clustering ensemble, SNNMF, ECMC*

Metric: accuracy

Experiments

Accuracy: among 11 tasks: 1 tie, 1 loss, 9 wins

Observed overfitting



Experiments

