

Why Ensemble/Boosting?

Make Weak Models Great Again

Zhedong Zheng



Outline

1. Intro
2. Random Ensemble
3. Strategic Ensemble

Collective Intelligence is not new

三个臭皮匠顶一个诸葛亮。

Two heads are better than one.

The wisdom of the crowd.

Background Diversity in University

Related Terminology

Mixture of Experts (MoE)

Multiple Agents

Model Ensemble

Self-boosting
Mutual Learning

Outline

1. Intro
2. Random Ensemble
3. Strategic Ensemble

yes

Celebrity Guess

Please start a guess celebrity game. I have an answer. You can ask me the question.



Great! Let's play a Guess the Celebrity game. I will try to guess the celebrity you are thinking of by asking yes or no questions.



No.



Is your chosen celebrity an actor or actress?



yes.



Does this musician or singer play an instrument professionally?



no.



Is this musician or singer primarily known for their work in the pop music genre?



no



Is this musician or singer primarily known for their work in the hip-hop or rap genre?



Is the band or group this musician is a part of known for blending different music genres in their work?

Given the previous answers, I'm trying to narrow down the possibilities. Since we've established that the celebrity is a singer who has been part of a rock band that is still active and has an international following, and considering the timeframe and genre-blending, could you tell me if the band you're thinking of is Coldplay? If not, does your band include elements of alternative rock in their music?



Yes. but not coldplay



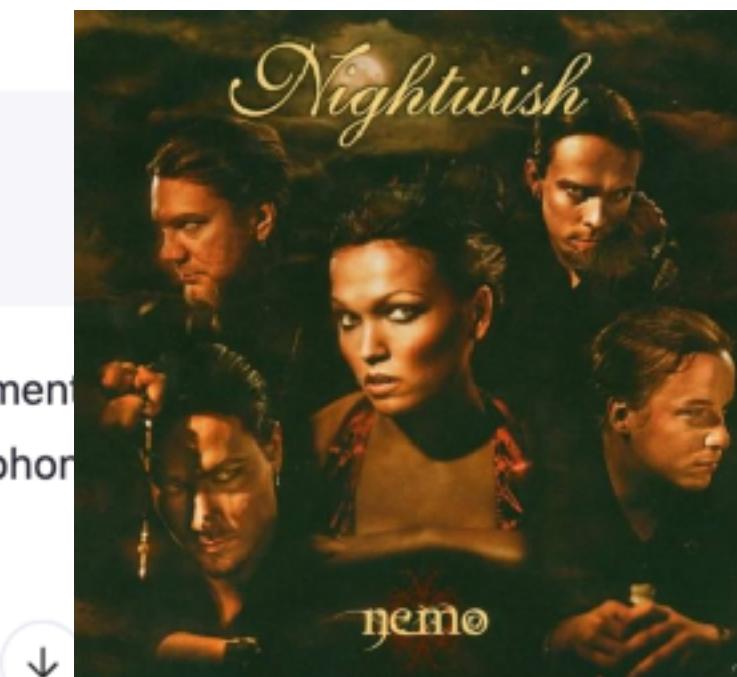
Got it, not Coldplay. Since we've established that the band includes elements of alternative rock and has an international following, could the band you're thinking of be Radiohead? If not, does this band have a style that leans more towards hard rock or progressive rock?



No. Metal.



Thanks for the additional clue. Given that the band plays metal and has an international following, could the band you're thinking of be Metallica? If not, does this band also incorporate elements of other genres into their music, such as folk or symphonic influences?



symphonic

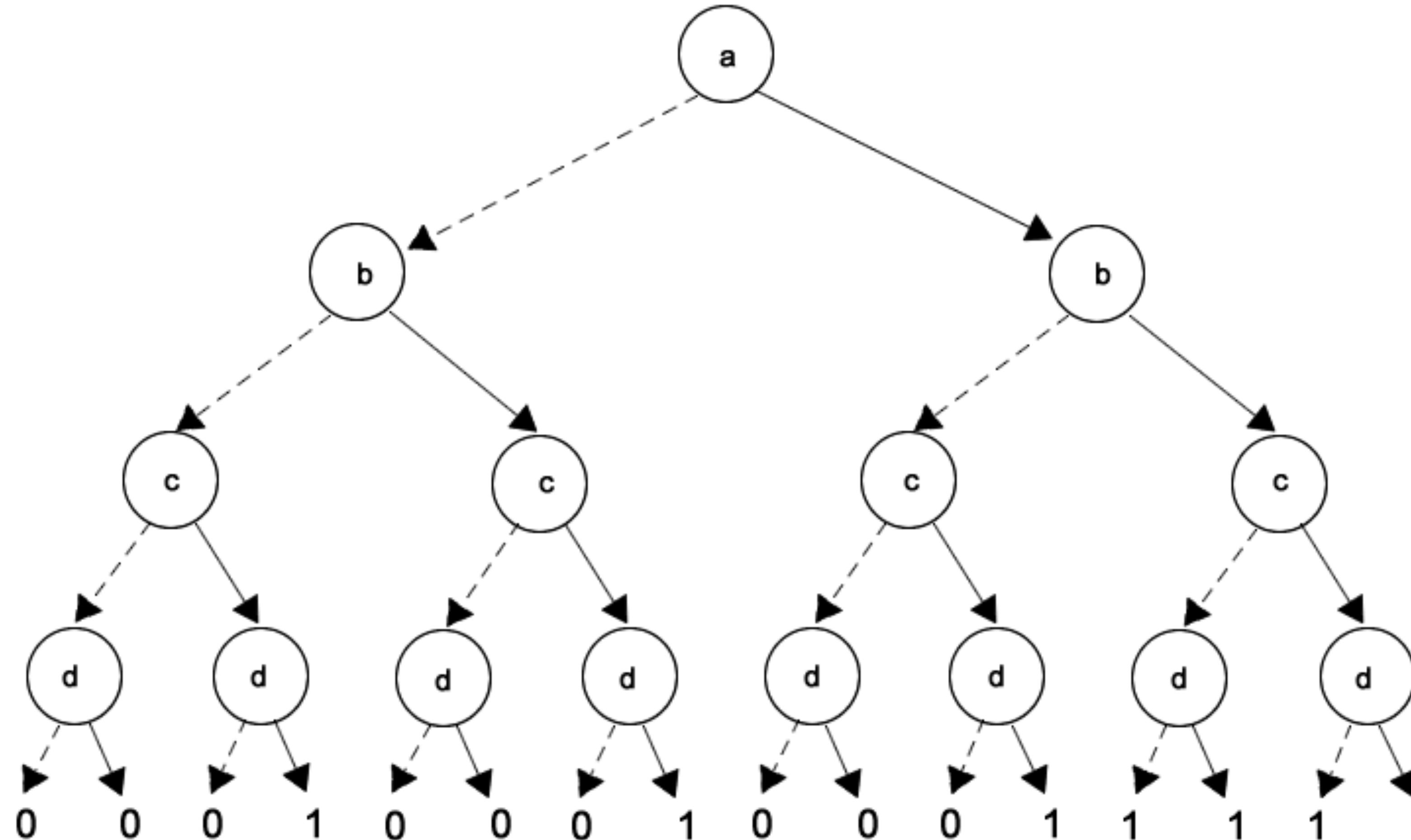


With the clue that the band incorporates symphonic elements, could the band you're thinking of be Nightwish? They are well-known for blending symphonic elements with metal and have a significant international following.

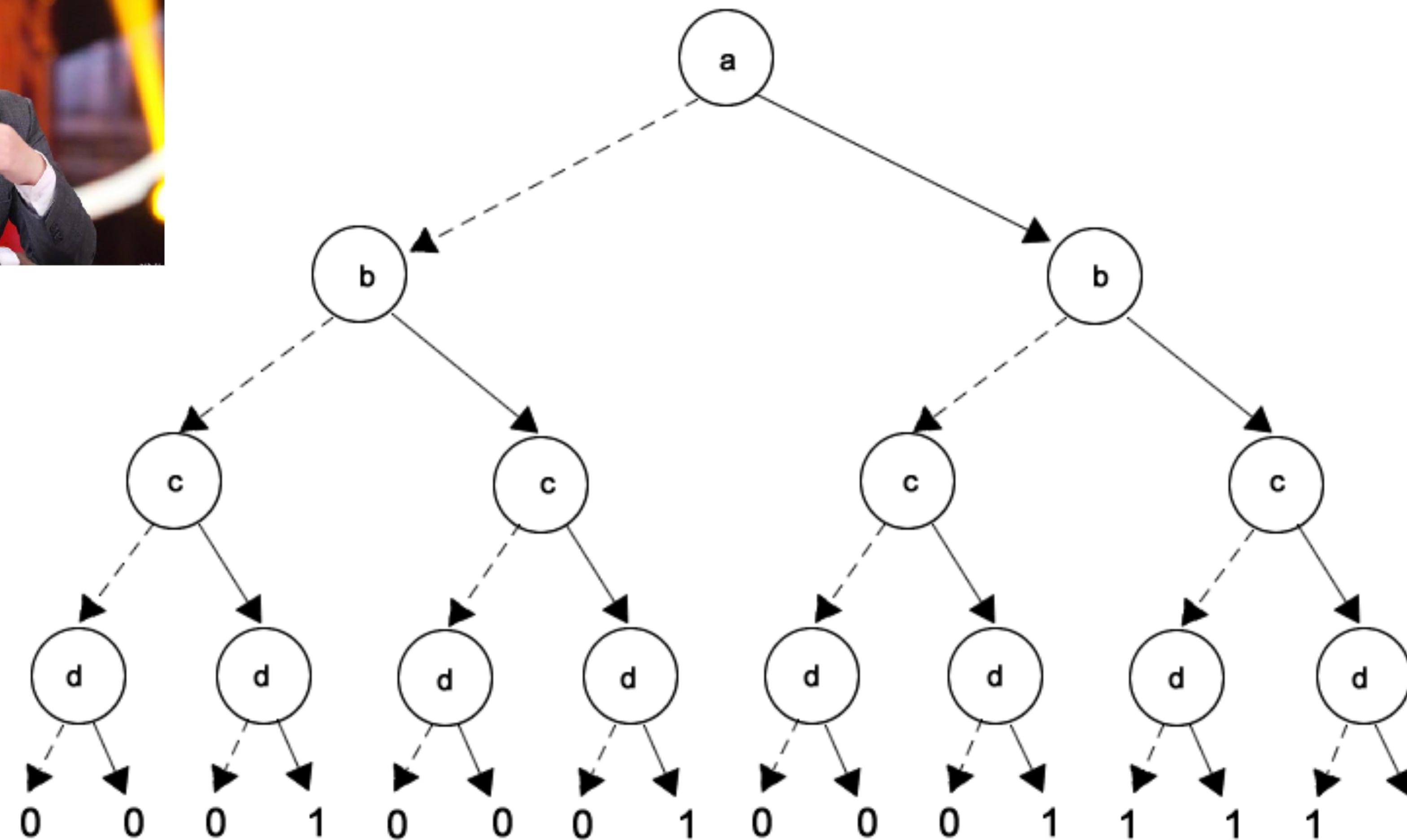


Decision Tree

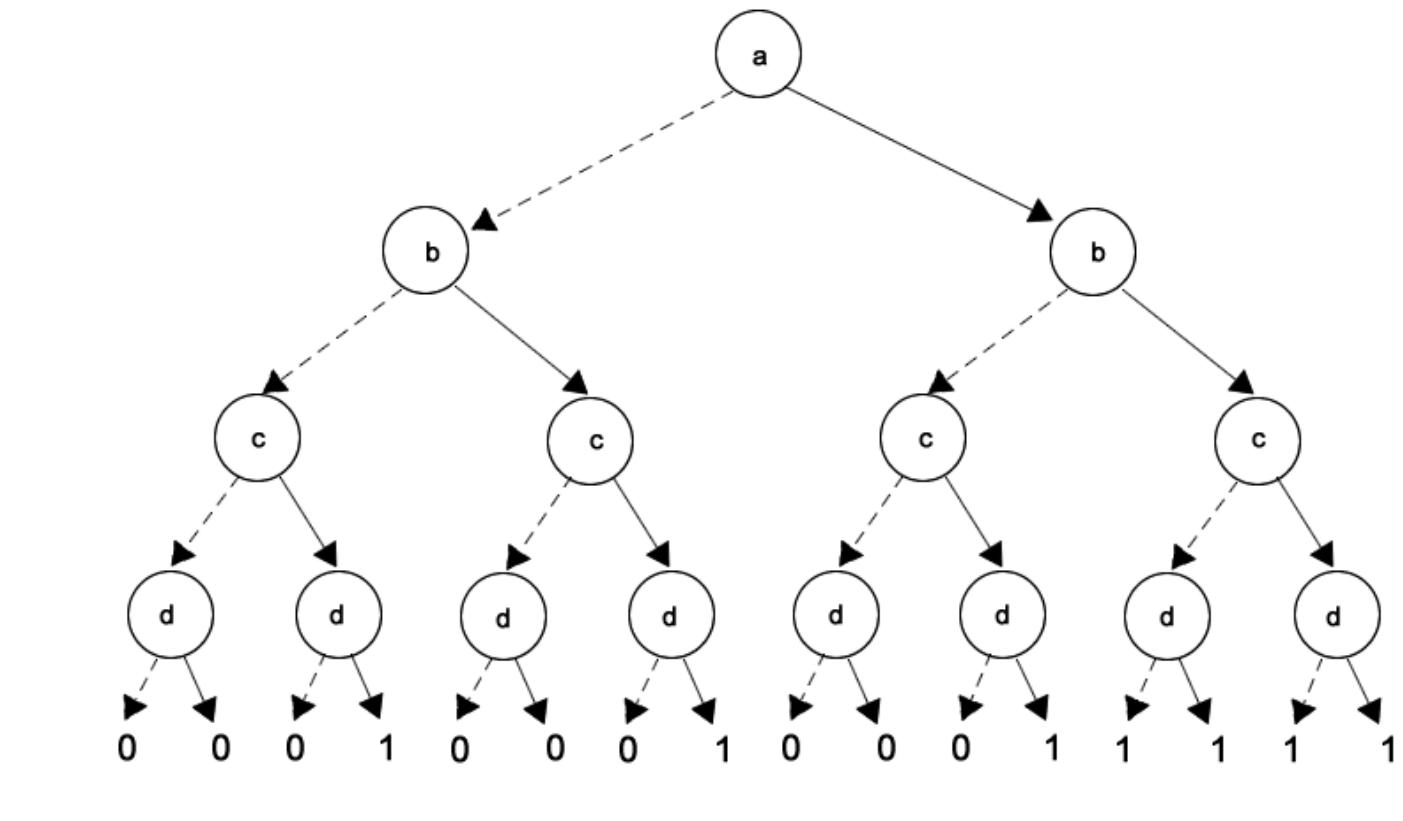
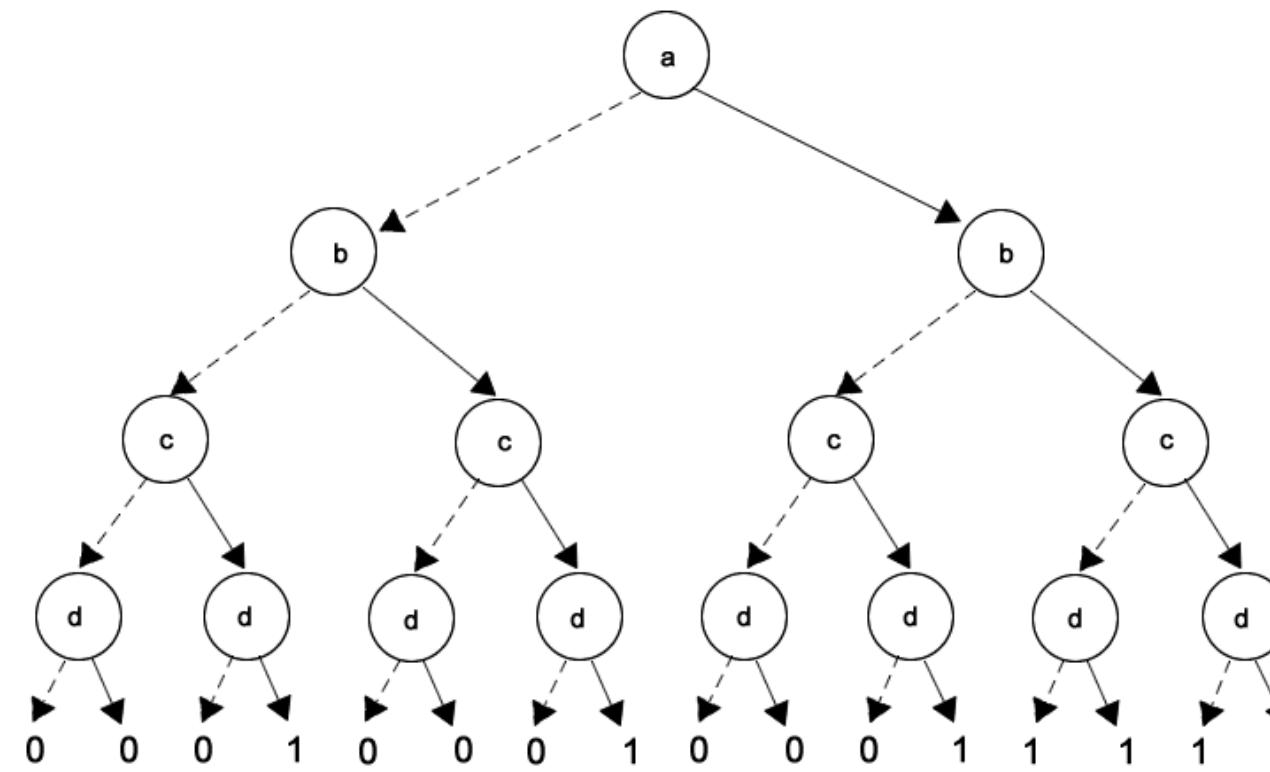
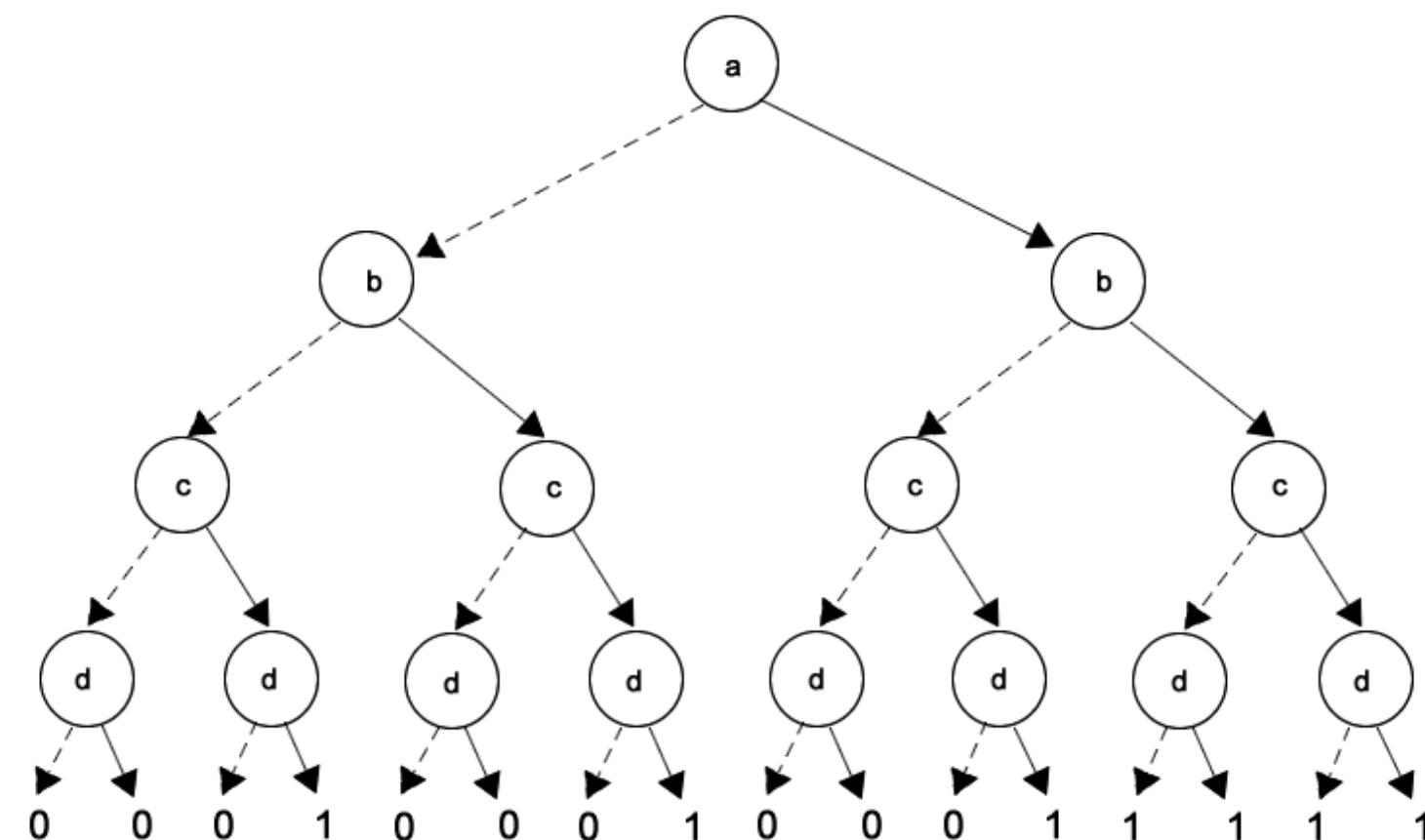
Every node is a binary classifier.



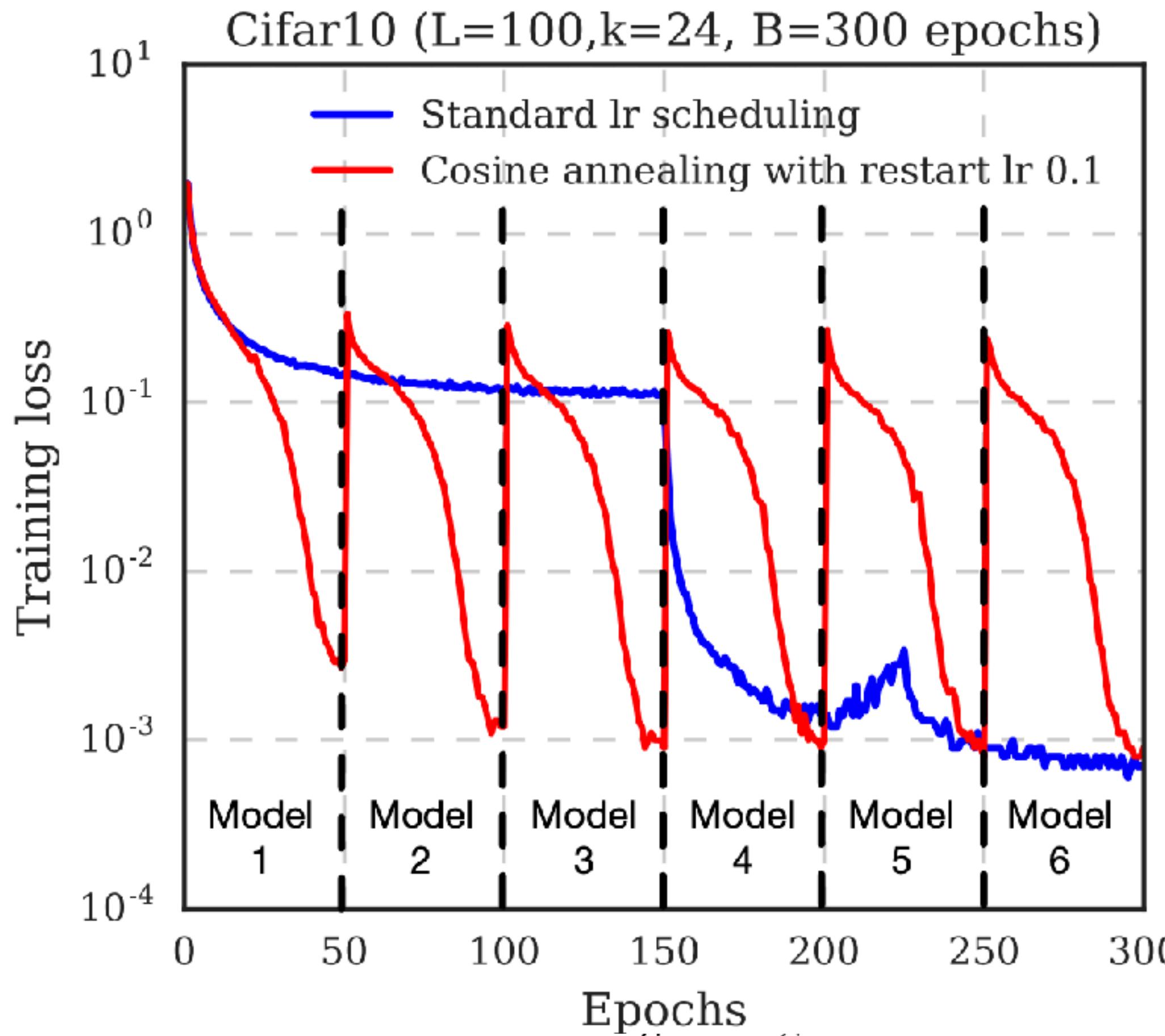
Meaningless question does not lead to answer



Decision Forest (Multiple Randomly-initialized Trees)



Transfer to Deeply-learned Model

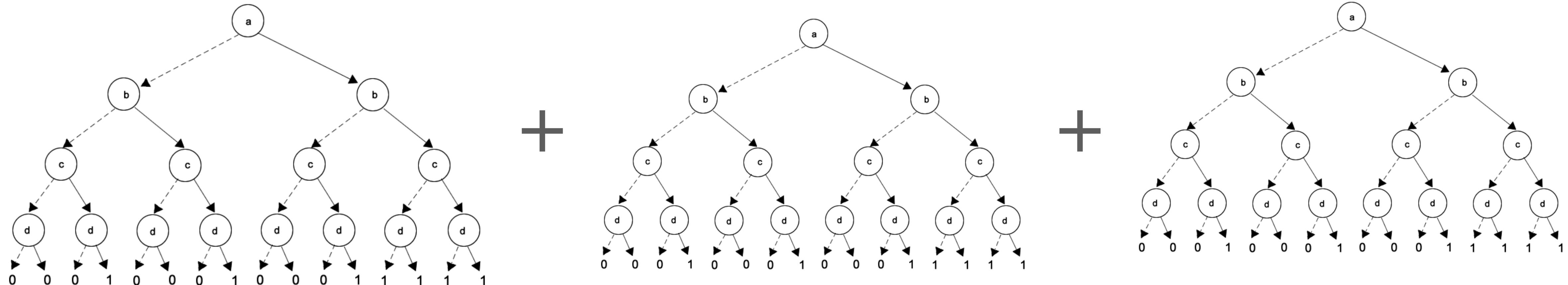


	Method	C10	C100	SVHN	Tiny ImageNet
ResNet-110	Single model	5.52	28.02	1.96	46.50
	NoCycle Snapshot Ensemble	5.49	26.97	1.78	43.69
	SingleCycle Ensembles	6.66	24.54	1.74	42.60
	Snapshot Ensemble ($\alpha_0 = 0.1$)	5.73	25.55	1.63	40.54
	Snapshot Ensemble ($\alpha_0 = 0.2$)	5.32	24.19	1.66	39.40
Wide-ResNet-32	Single model	5.43	23.55	1.90	39.63
	Dropout	4.68	22.82	1.81	36.58
	NoCycle Snapshot Ensemble	5.18	22.81	1.81	38.64
	SingleCycle Ensembles	5.95	21.38	1.65	35.53
	Snapshot Ensemble ($\alpha_0 = 0.1$)	4.41	21.26	1.64	35.45
DenseNet-40	Snapshot Ensemble ($\alpha_0 = 0.2$)	4.73	21.56	1.51	32.90
	Single model	5.24*	24.42*	1.77	39.09
	Dropout	6.08	25.79	1.79*	39.68
	NoCycle Snapshot Ensemble	5.20	24.63	1.80	38.51
	SingleCycle Ensembles	5.43	22.51	1.87	38.00
DenseNet-100	Snapshot Ensemble ($\alpha_0 = 0.1$)	4.99	23.34	1.64	37.25
	Snapshot Ensemble ($\alpha_0 = 0.2$)	4.84	21.93	1.73	36.61
	Single model	3.74*	19.25*	-	-
	Dropout	3.65	18.77	-	-
	NoCycle Snapshot Ensemble	3.80	19.30	-	-
	SingleCycle Ensembles	4.52	18.38	-	-
	Snapshot Ensemble ($\alpha_0 = 0.1$)	3.57	18.12	-	-
	Snapshot Ensemble ($\alpha_0 = 0.2$)	3.44	17.41	-	-

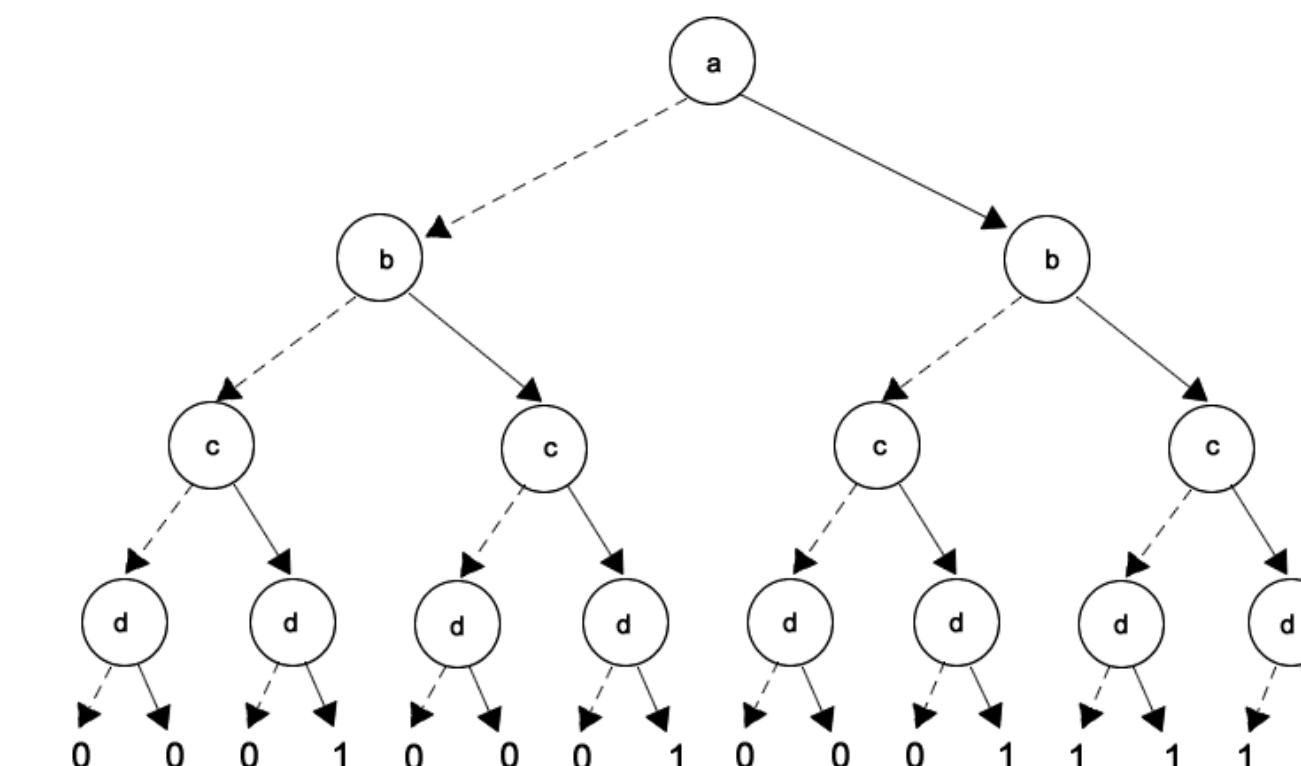
Ensembling at Test Time. The ensemble prediction at test time is the average of the last m ($m \leq M$) model's softmax outputs. Let \mathbf{x} be a test sample and let $h_i(\mathbf{x})$ be the softmax score of snapshot i . The output of the ensemble is a simple average of the last m models: $h_{\text{Ensemble}} = \frac{1}{m} \sum_0^{m-1} h_{M-i}(\mathbf{x})$. We always ensemble the last m models, as these models tend to have the lowest test error.

Snapshot Ensembles: Train 1, get M for free. ICLR, 2017

Temporal Ensembling = One Final Model.



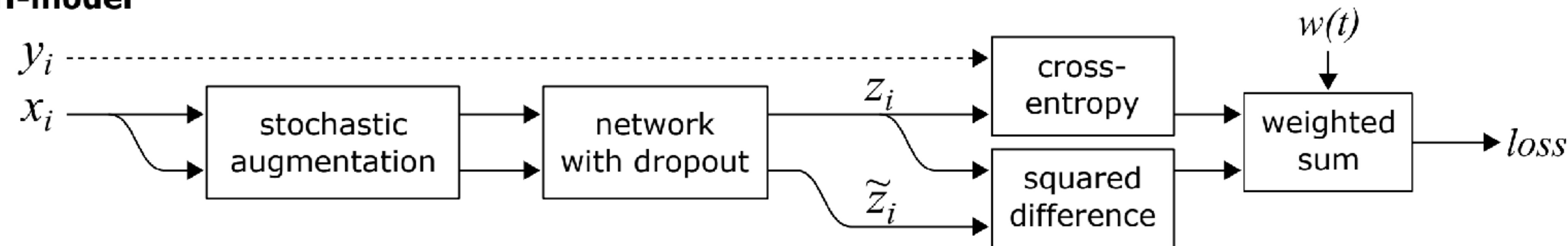
History Distillation



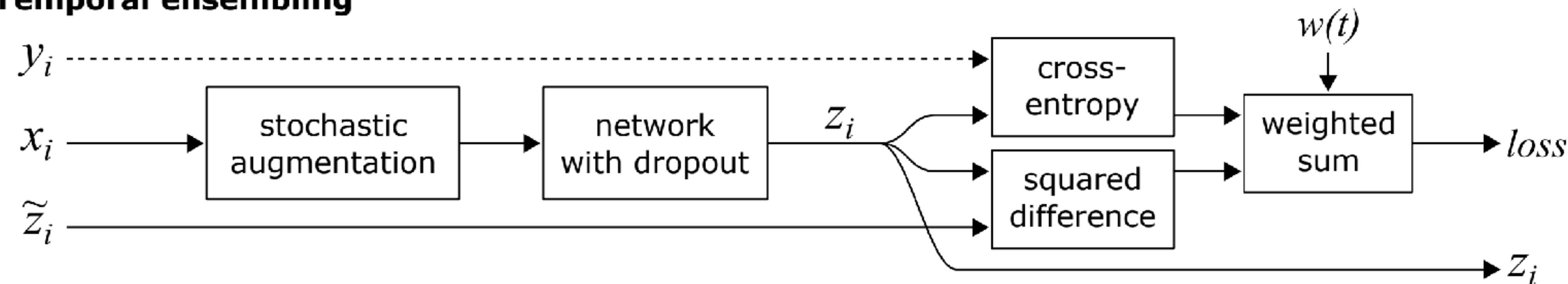
Temporal Ensembling for Semi-Supervised Learning ICLR 2017

Temporal Ensembling <- History Representation

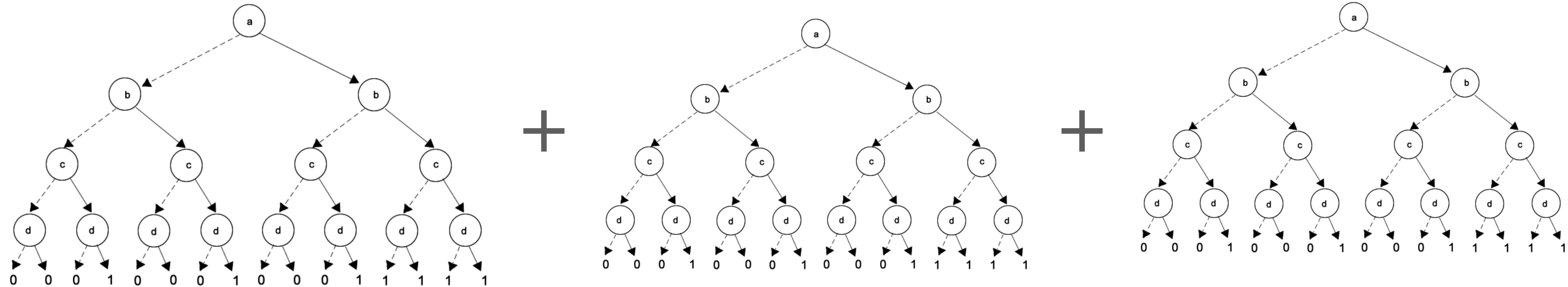
Π -model



Temporal ensembling

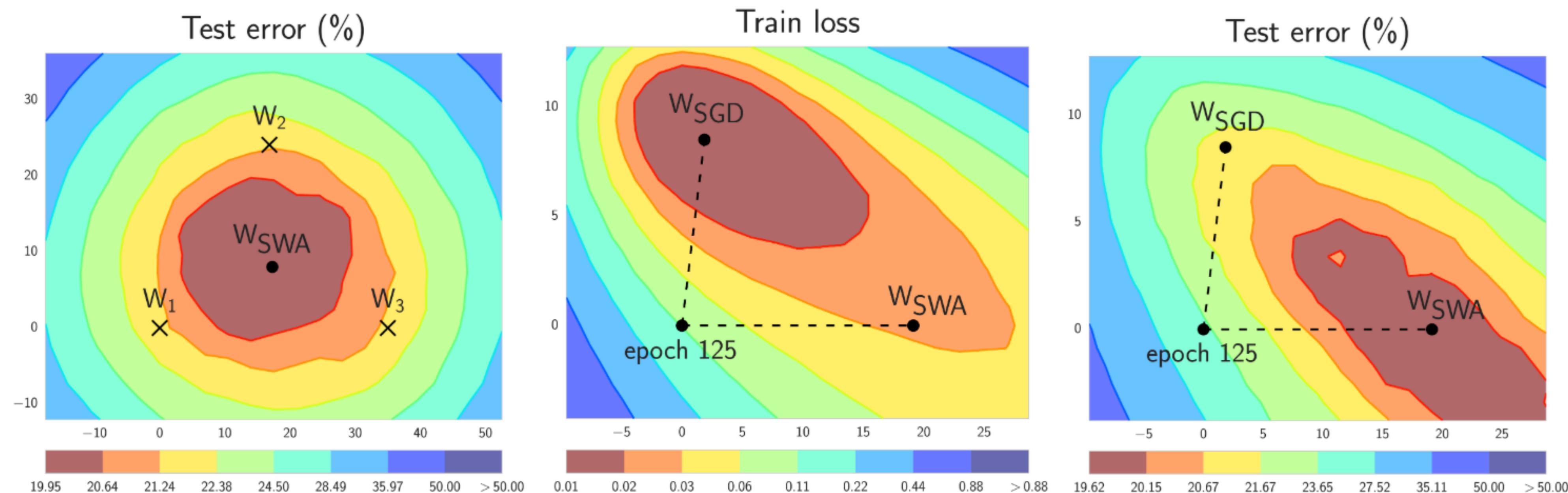


MeanTeacher = Temporal Fusion (History Weight Version / SWA Version)



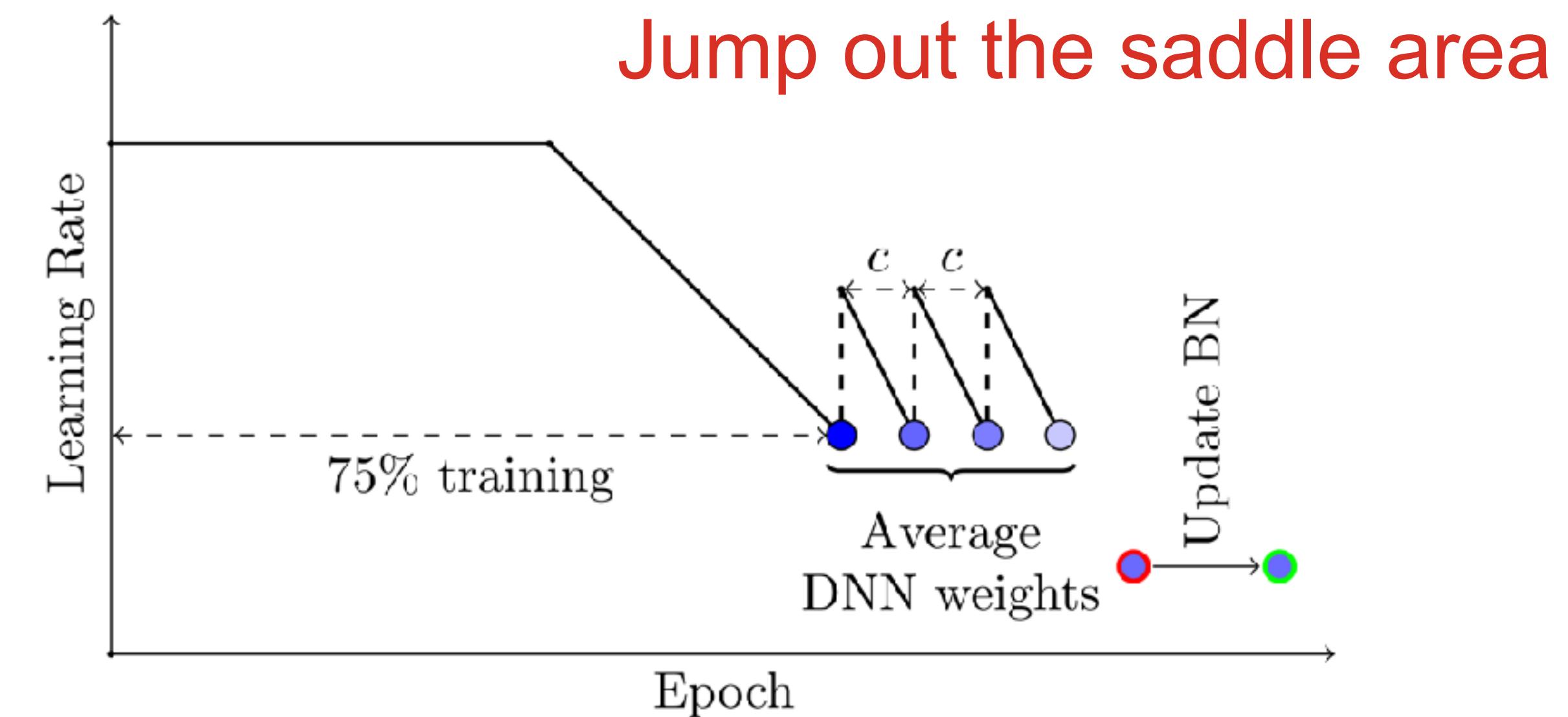
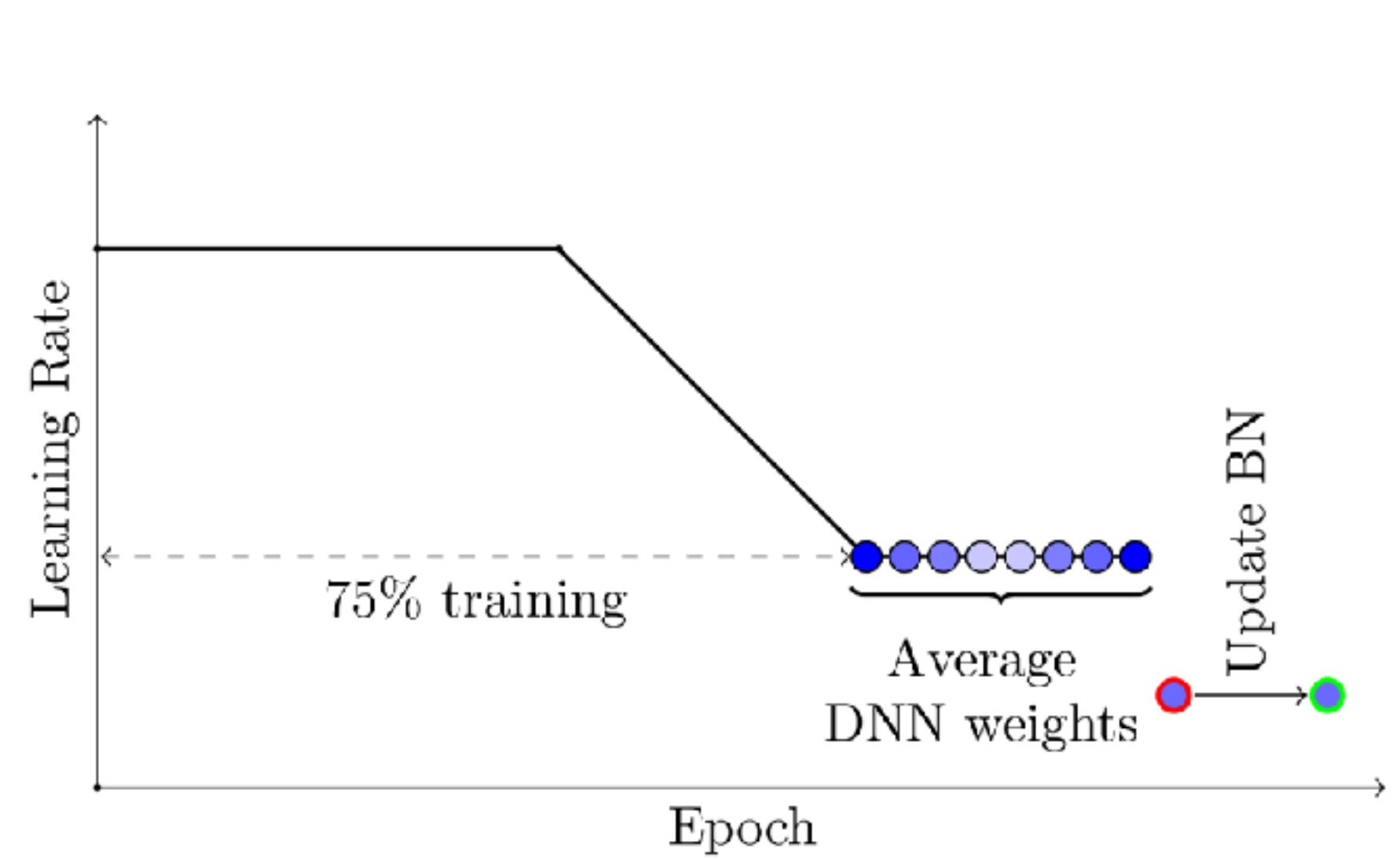
Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results NeurIPS 2017
<https://pytorch.org/blog/stochastic-weight-averaging-in-pytorch/>

MeanTeacher = Temporal Fusion (History Weight Version / SWA Version)



Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results NeurIPS 2017
<https://pytorch.org/blog/stochastic-weight-averaging-in-pytorch/>

MeanTeacher = Temporal Fusion (History Weight Version / SWA Version)



Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results NeurIPS 2017
<https://pytorch.org/blog/stochastic-weight-averaging-in-pytorch/>

Outline

- 1. Intro**
- 2. Random Ensemble**
- 3. Strategic Ensemble**

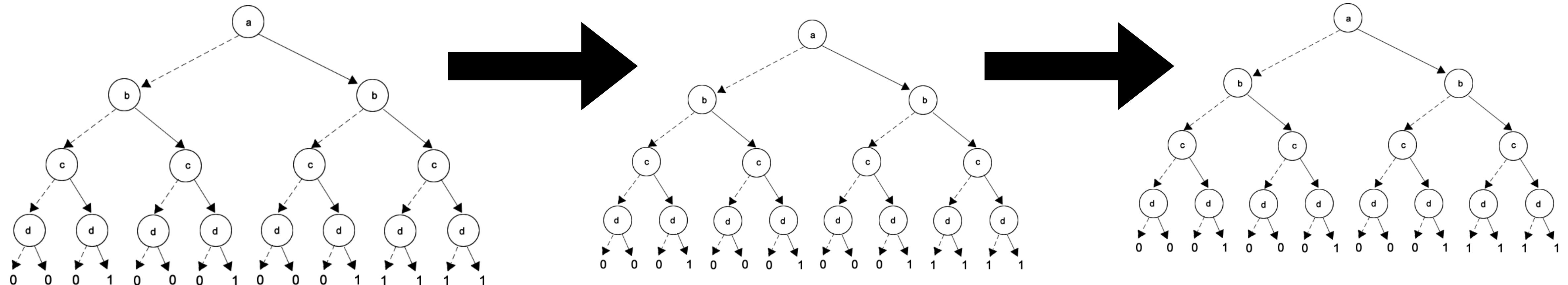
Key Question

Diversity and Complementary

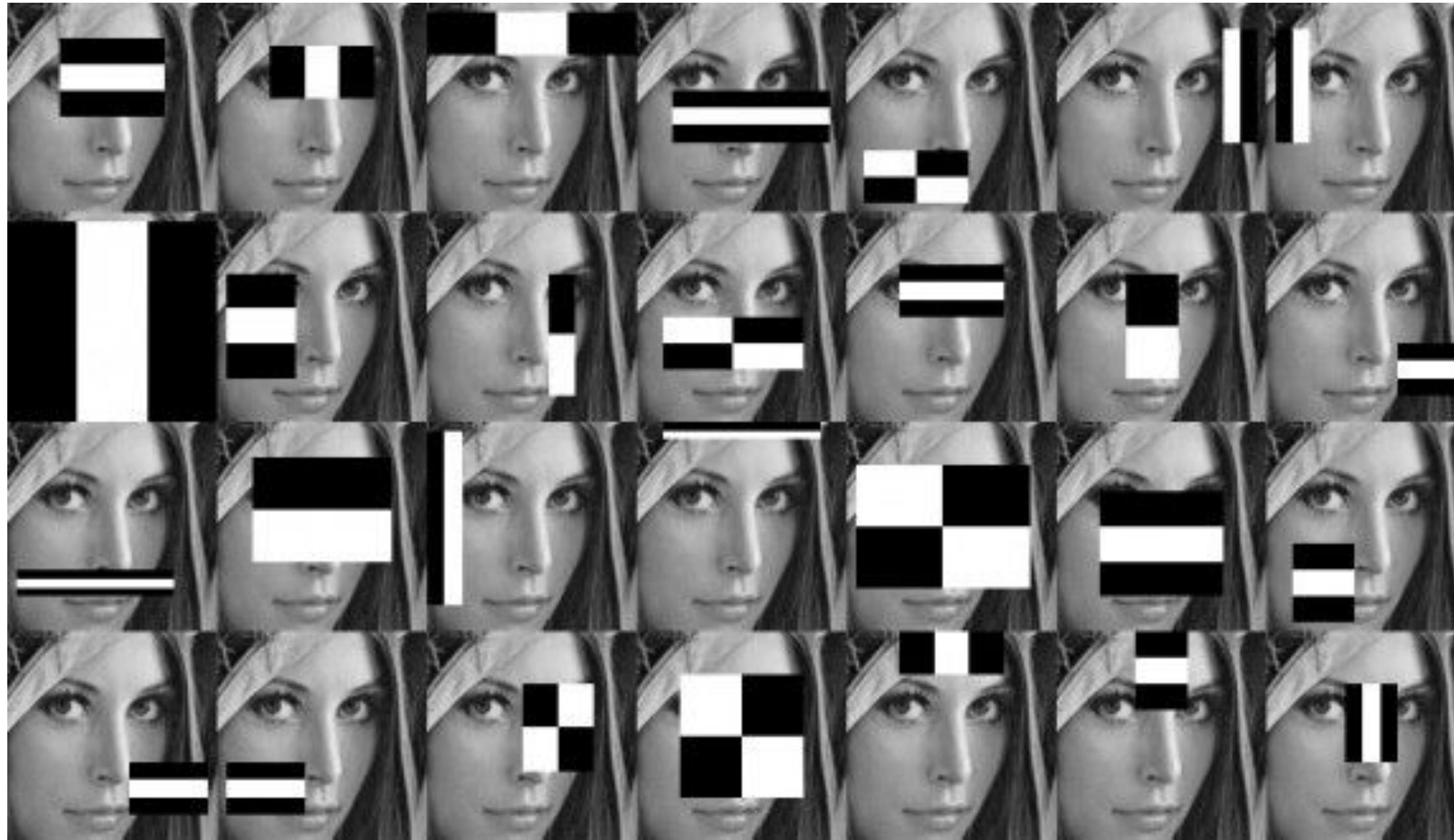


Adaboost = A Better Decision Forest

(Next Tree is trained on the former failures.)



Face Detector (Have/No Faces) Viola&Jones



If no face,
return; else
go to next
classifier.

$$h(\mathbf{x}) = \text{sgn} \left(\sum_{j=1}^M \alpha_j h_j(\mathbf{x}) \right)$$

Face Detector (Have/No Faces) Deep learning

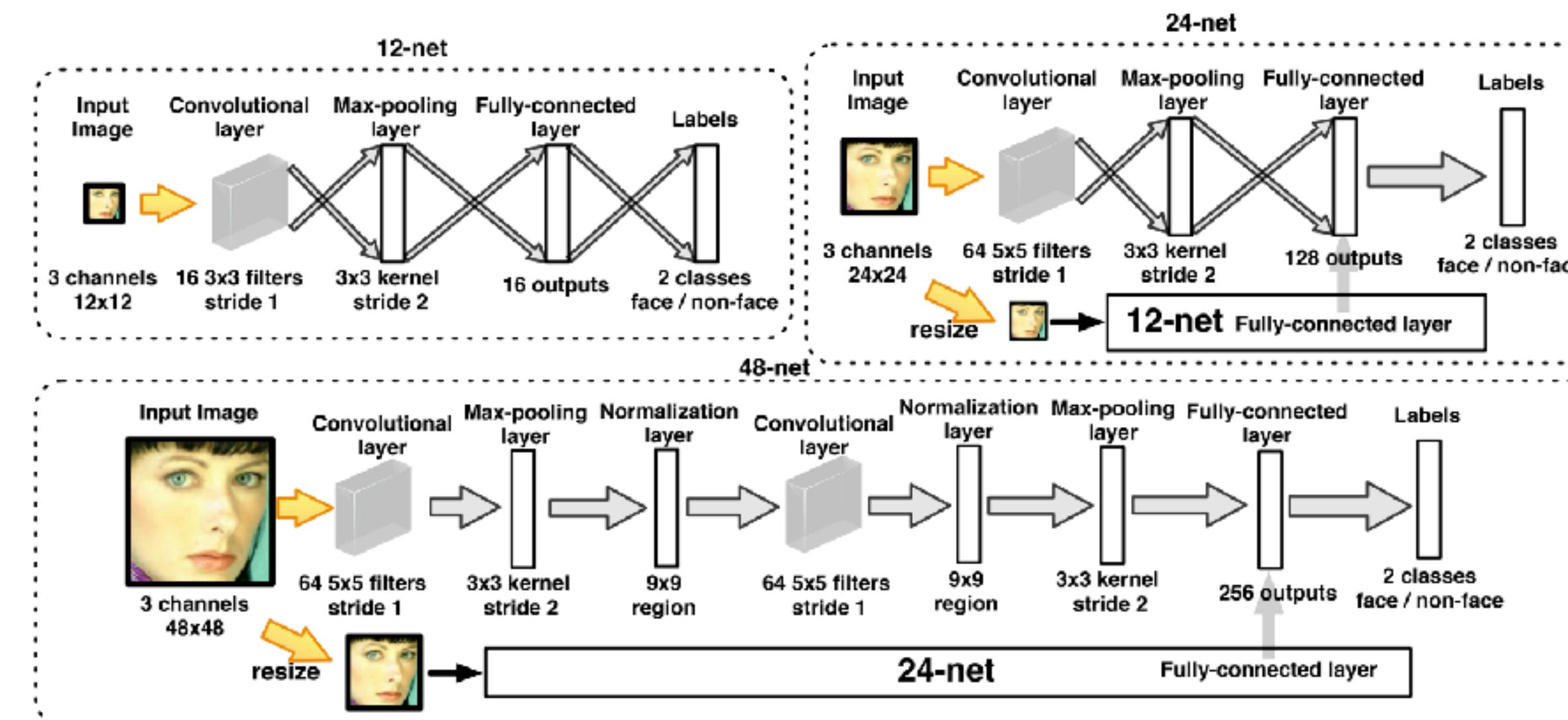
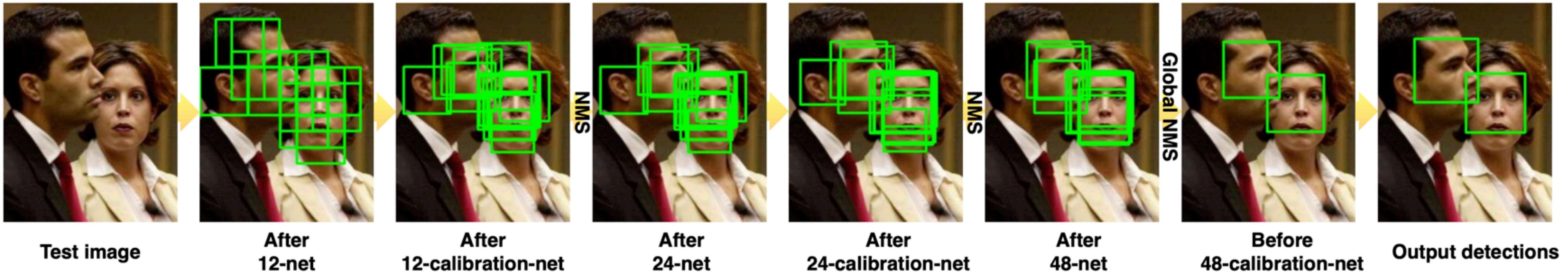


Figure 2: CNN structures of the 12-net, 24-net and 48-net.

Adaboost Demonstration

