What is the Effect of Importance Weighting in Deep Learning?

Jonathon Byrd and Zachary C Lipton
Carnegie Mellon University

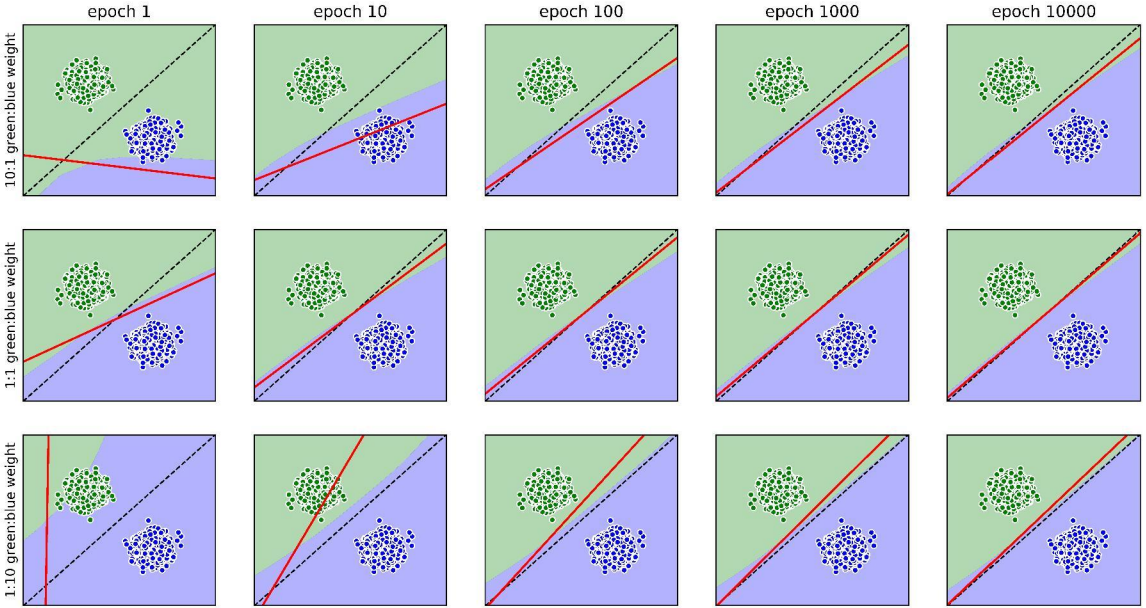
Email: jabyrd@cmu.edu
Paper: https://arxiv.org/abs/1812.03372
Presentation: Pacific Ballroom #19, 6:30–9:00 PM

Importance Weighting

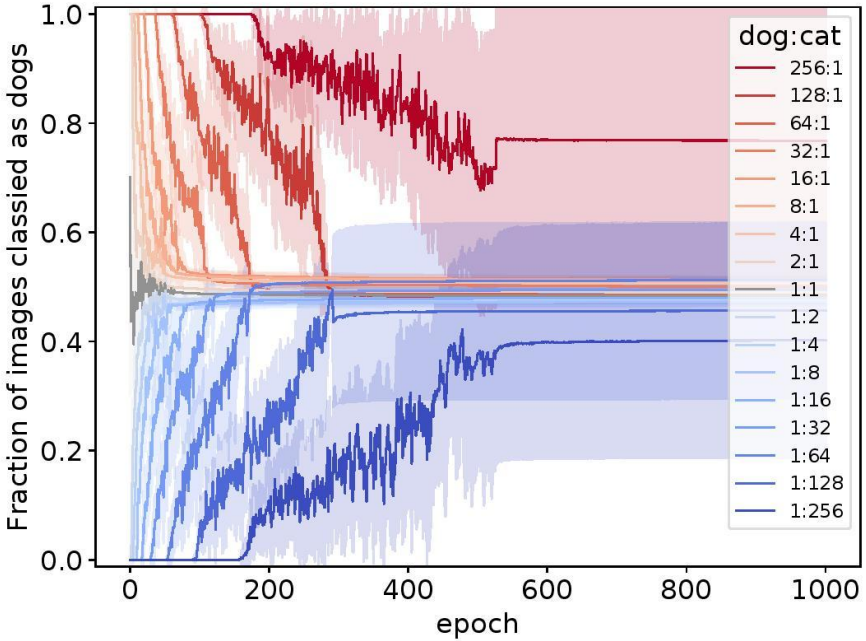
- Importance weighting (IW) is a key element in many ML algorithms
- It's behavior is well understood with simple, misspecified models
- Recent theory shows that (linear) neural networks, trained on separable data converge to the (weight-invariant) max-margin solution (Soudry et al 2017)
- Only the location of the data determines the solution, not their density/weight
- Many papers use IW-ERM with deep learning models, **for which training sets of practical interest are often separable**

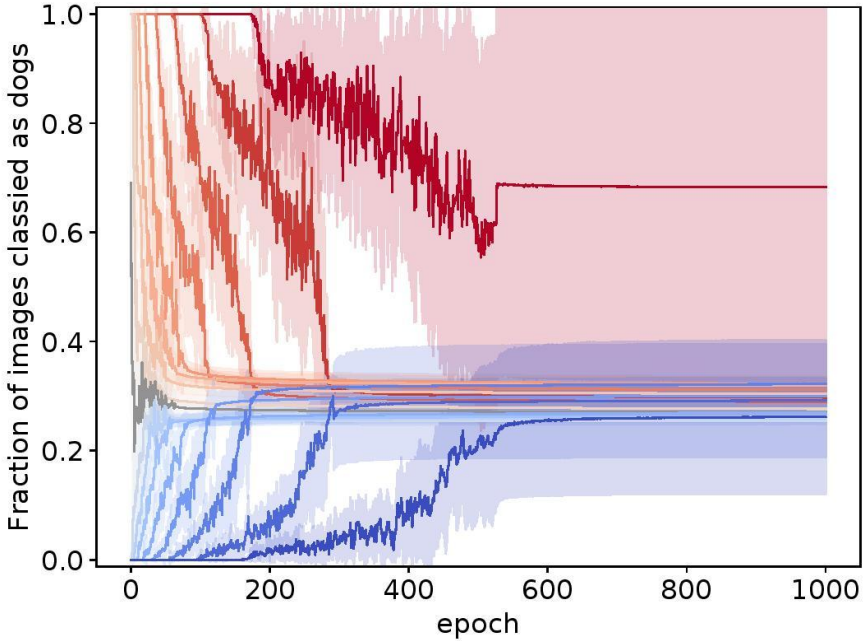
Toy Example

- 2D synthetic data
- 1-layer MLP
- Effect of IW diminishes Over training iterations

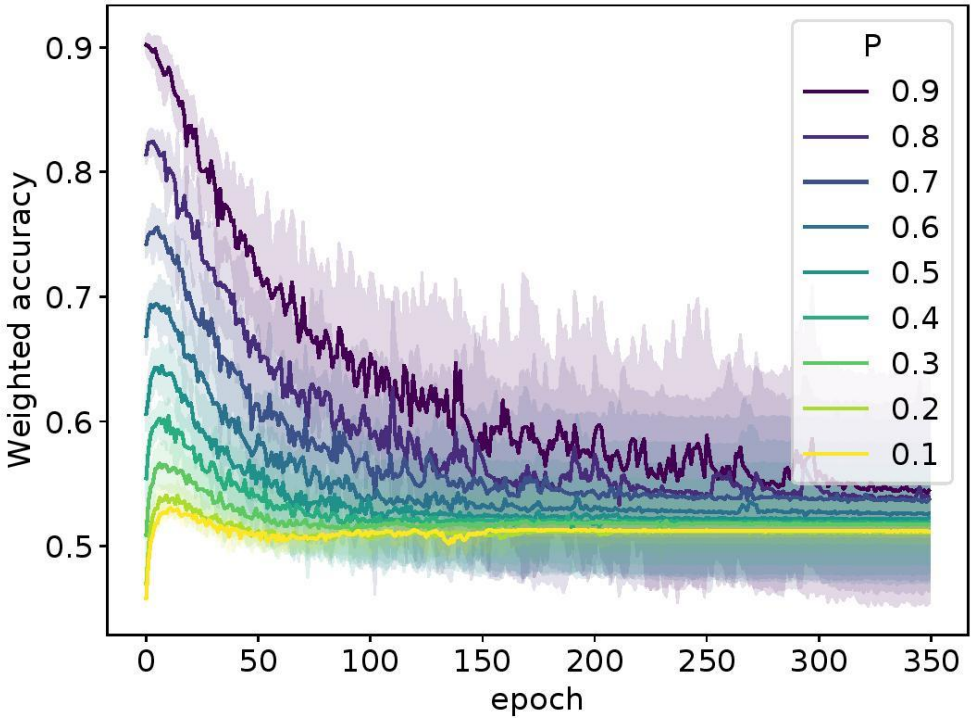


CIFAR–{cats, dogs} IW depends on early stopping

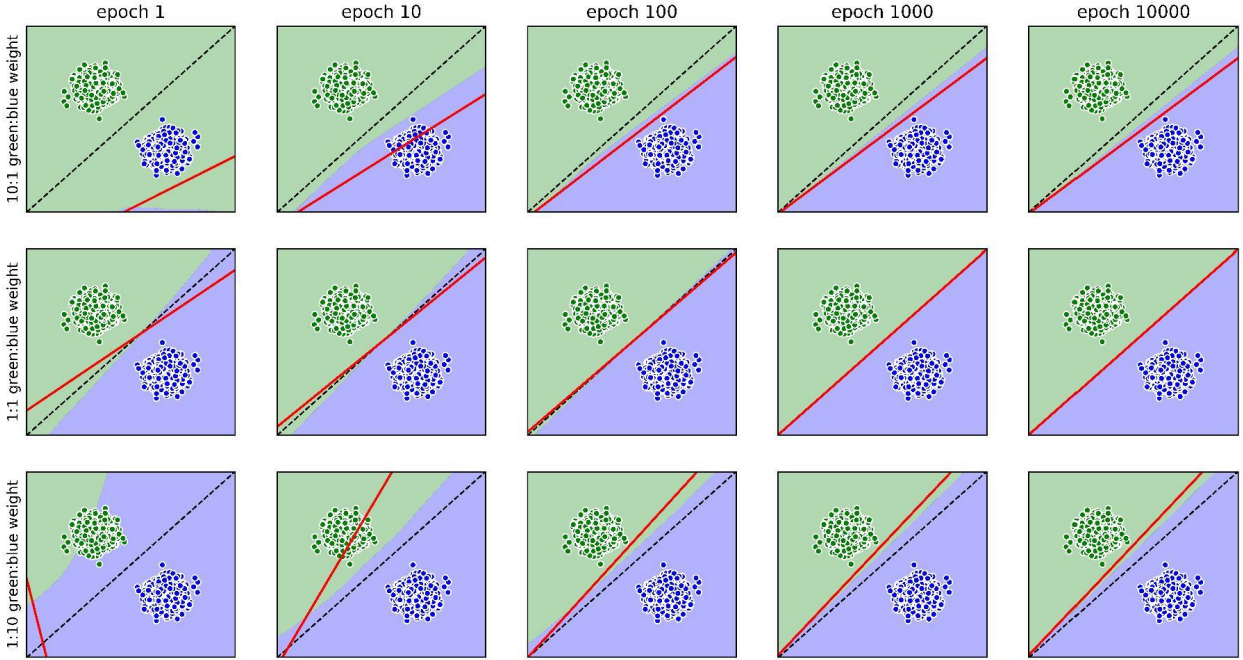




Usefulness of IW to adjust for label shift depends on early stopping



The effects of importance weighting depend on regularization. (L2-regularization shown)



Thank you

Thu Jun 13th 06:30 -- 09:00 PM
Pacific Ballroom #19