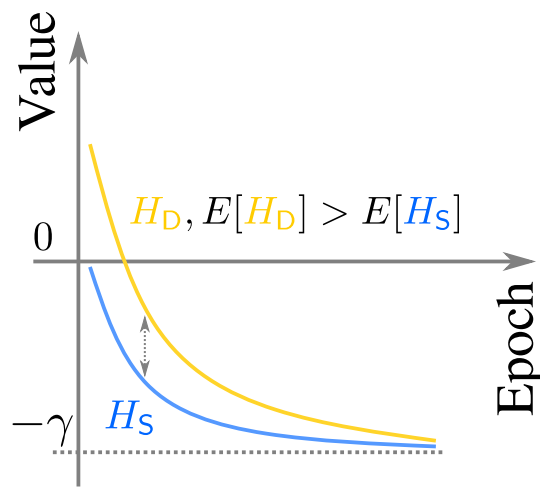
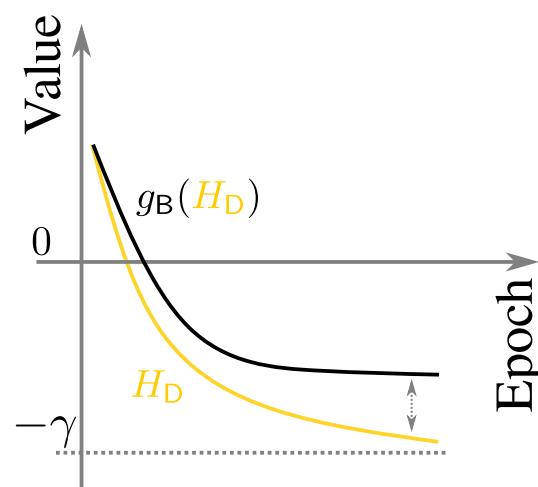


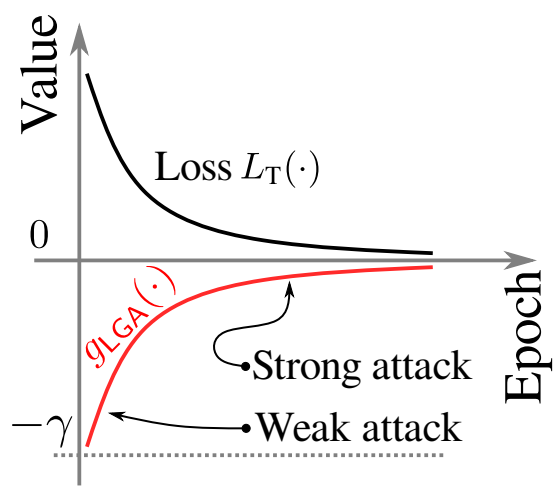
(a) Relationship between loss and hardness when $L_T(\cdot) > 0$



(b) Shrunk $E[H_D - H_S]$ value leads to weak attack



(c) Slightly Boosting H_D for slightly stronger attack



(d) Linear Gradual Adversary for balanced training