

When performing cross validation, do you only look at the raw average validation scores, or do you also look at the training scores and try and minimize the generalization gap? Previously, I only looked at the former, but recent experiments yielded a much lower generalization gap and this has me questioning whether I should be optimizing for this as well.