

surajp:

I wasn't a big fan of ensembles in production (but turns out pruning and distillation can improve generalization. Is it true [@jrb?](#))

I don't quite understand. What do model pruning and model distillation have to do with ensembling? Unless of course, if you're talking about using model distillation to train a smaller student model using an ensemble of larger models as the teacher model.