Proj B tips

- See pinned post on piazza

- Common issues:
  - If you use GridSearchCV or RandomSearchCV with a custom splitter for Prob 1
    - Avoid **n_jobs = -1 : not reproducible!**
    - Be sure you use "refit=False"
      - Otherwise, will automatically "refit" on entire dataset (train+validation)
      - You might get balanced acc that is "too good" (>0.9)

Proj B tips

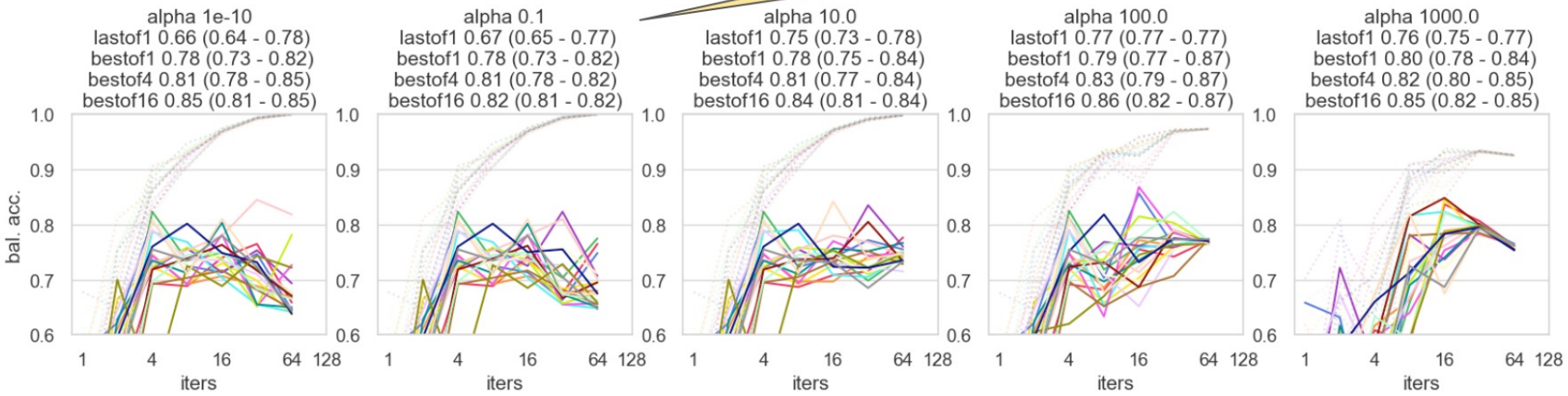
- What to submit to the leaderboard?
  - The best-performing model you found
    - e.g. in Problem 1, if your 1B model was better than 1D, use that
    - Be clear about this in your report!
  - If in Problem 2, you tried something and it didn't work well
    - Submit the "simpler" model that did better
    - Write about "what didn't work" so we know you tried it

Problem 1E: *duplicated* train data

(dup rare class up to size of biggest class)

Fit MLP with 100 units, L-BFGS solver

0.67 BA with no early stopping
0.78 with early stop
0.81 with early stop + best of 4 rand seeds
0.82 with early stop + best of 16 rand seeds



Most runs overfit after 4 iters

Most runs level off

Enough regularization to not be perfect at training, we still see decent performance

Random initialization matters! (so does early stopping)

Leaderboard

Rank	Submission Name	bal_acc	frac_data	acc_dres	acc_pull	acc_top
1	team zerrin yalan dunya	0.9287	0.5	1	0.9	0.8182
2	jk classify	0.9274	0.5	0.9455	0.9	0.8636
3	Siara and Bill	0.9126	0.5	0.9273	0.925	0.8409
4	Alex and Etai	0.9106	0.5	0.9636	0.85	0.7955
5	Sam & Charlotte	0.9067	0.5	1	0.9	0.7273
6	Nicole and Nathan	0.9054	0.5	1	0.9	0.75

Leaderboard

Rank	Submission Name	bal_acc	frac_data	acc_dres	acc_pull	acc_top	acc_trou
1	🙏	0.96	0.5	0.9818	0.975	0.8409	0.9804
2	Charlotte + Sam	0.9521	0.5	0.9636	0.9	0.8864	0.9804
3	Ash & Khaled	0.9517	0.5	0.9818	0.9	0.8864	0.9608
4	Zack	0.9501	0.5	1	0.875	0.8636	0.9804
5	JH_WCM	0.9496	0.5	1	0.925	0.8636	0.9804
6	test1 attempt 2	0.9477	0.5	1	0.95	0.8636	0.9804

$$E_{\text{hold-out}} = E_{\text{train}} + \underbrace{(E_{\text{hold-out}} - E_{\text{train}})}_{\approx \text{ generalisation gap}}$$

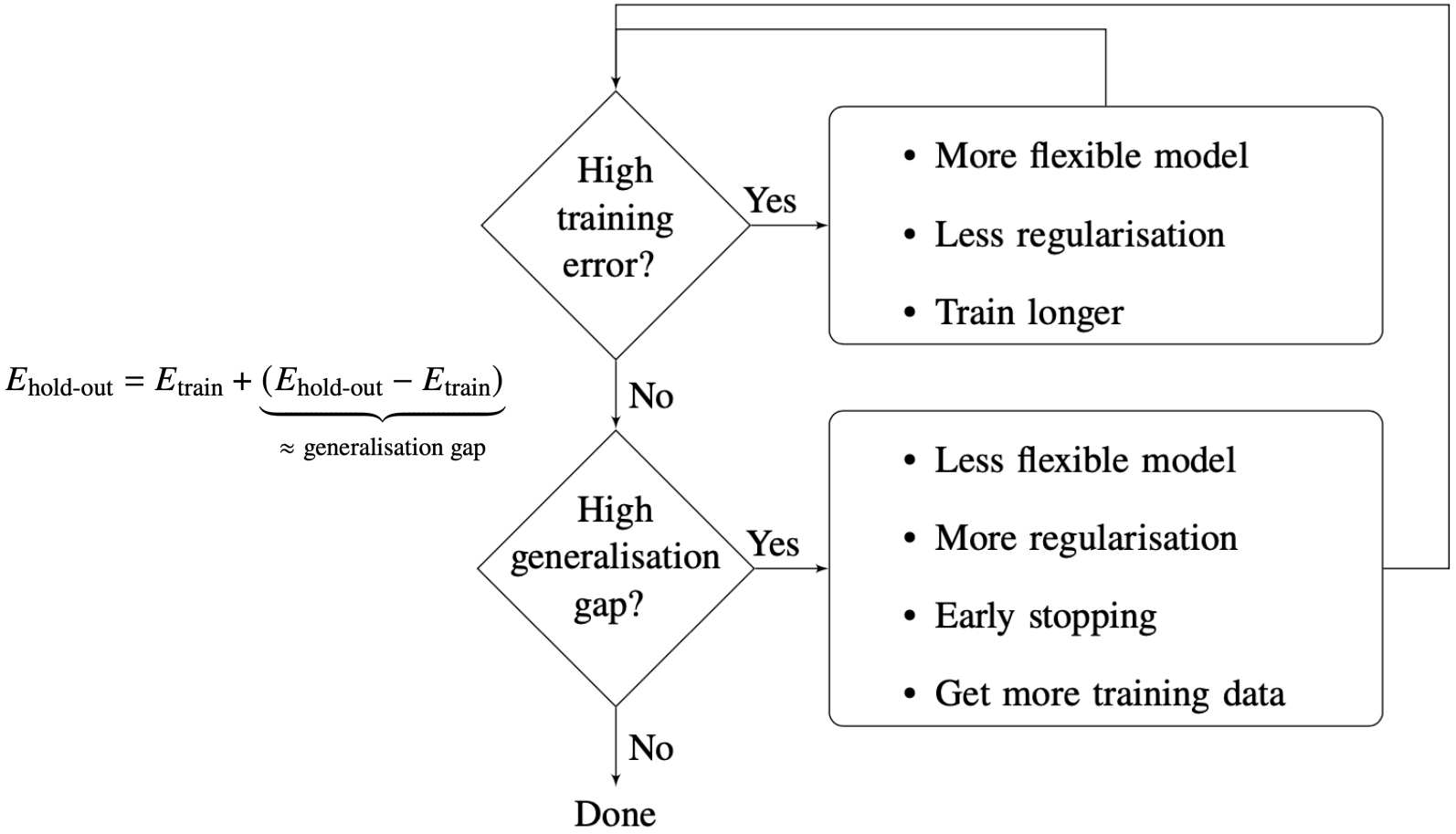


Figure 11.2: The iterative procedure of improving a model based on the decomposition of the validation error into the training error and generalisation gap.

When will adding data help?

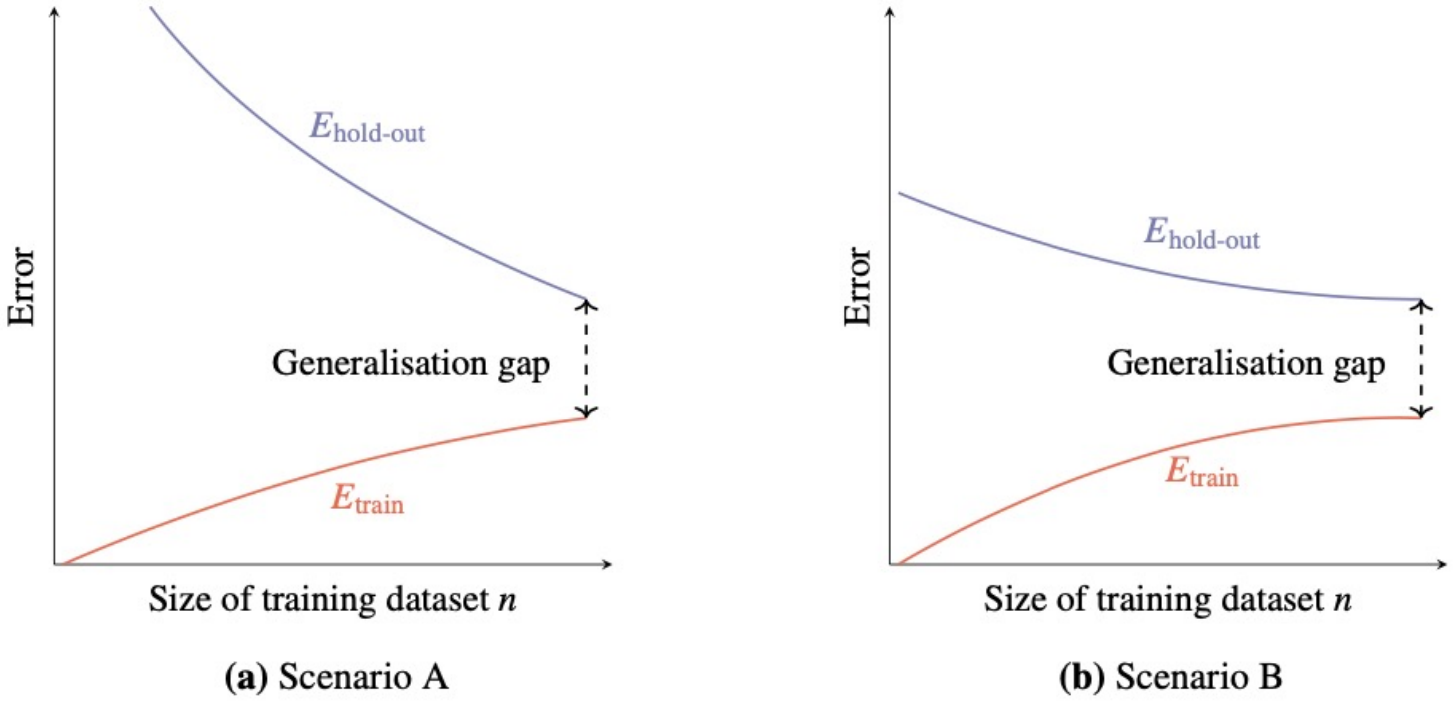


Figure 11.3: Learning curve for two different scenarios. In Scenario A, we can expect an improvement in the generalisation gap by collecting more training data, whereas in Scenario B, we are less likely to see an immediate improvement by adding more data.