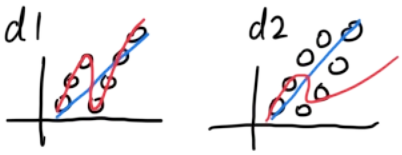
8) Overfitting

Overfitting

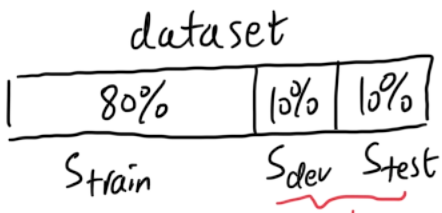
/ = 2-param model
/ = 4-param model



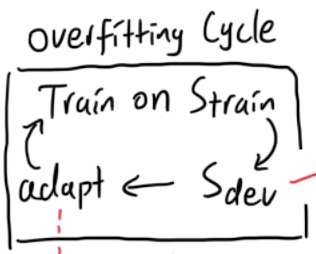
- Situation when the model fits the dataset perfectly
- Problem: fails to generalize data ; may perform bad on new datasets
- Solution: ① Cross-Validation ② Regularization

Cross Validation

- Process:



⇒



- Assumption : Each set is independent & uncorrelated of each other

Generalization

- Situation that model works well when applied to new data
- Once determine generalization boundary (population), Cross-Validate it

larger boundary = larger error

Batch

- # samples the model trains on at a time during an epoch
- Process : Strain

⇒ ith batch_acc = $\frac{\#\text{ correct}}{\text{batch\_size}}$

training_acc = mean (batch_acc)

test_acc = # correct / total

- Using this wisely can boost model learning speed