

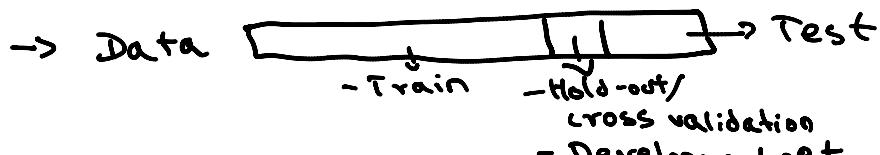
Week 1a - Setting up your Machine Learning Application

Friday, August 14, 2020 11:24 PM

① Train / Dev / Test Sets

→ Applied ML is a highly iterative process

#layers, #hidden units, learning rates, activation fns... etc



In previous times: 70/30%. (or) 60/30/10%.
100 - 1000 - 10000 → training examples

Big data: 1,000,000+ examples

98% / 1% / 1%
or even higher train sets

→ Mismatched train/test dist.

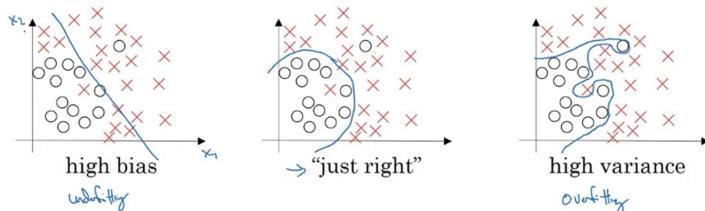
Training set:
pictures from
webpages
↑
usually Higher
quality

Test set:
pictures from
users of app

- Ensure same distribution among all sets
- Not having test set might be fine (Only dev set)
↳ used to iterate with diff. models

② Bias / Variance

Bias and Variance



eg: if Train set error
 Dev set error
 Human $\approx 0\%$.
 \Rightarrow Optimal (Bayes) error
 $= 0\%$.

| | 1%. | 15%. | 15%. | 0.5%. |
|-------------------------------------------------|-----|------|------|-------|
| high variance | 11% | 16% | 30% | 1% |
| under fitting | | | | |
| (high bias) | | | | |
| high variance | | | | |
| (high bias) | | | | |
| ↓ | | | | |
| maybe a good model if bayes error was higher | | | | |

③ Basic Recipe for ML

