



deeplearning.ai

Mismatched training and dev/test data

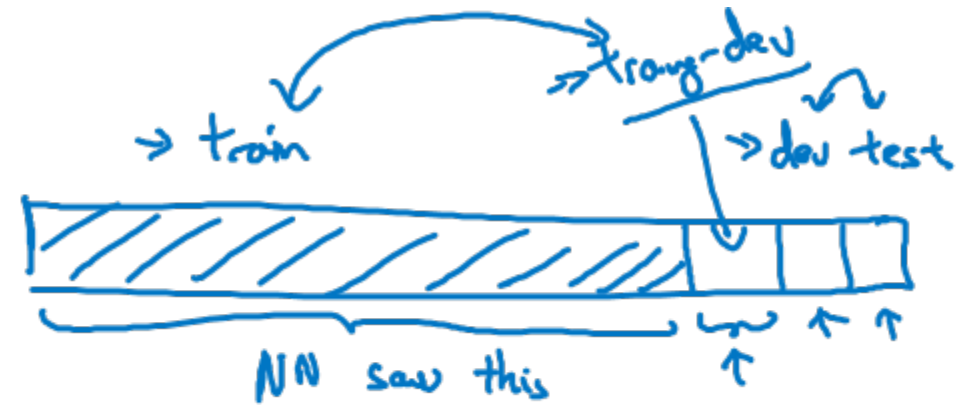
Bias and Variance with mismatched data distributions

Cat classifier example

Assume humans get $\approx 0\%$ error.

Training error 1%
 Dev error 10% $\downarrow 9\%$

Training-dev set: Same distribution as training set, but not used for training



Training error	1%		1%
→ Training-dev error	9%	↑ variance	1.5%
→ Dev error	10%		10% ↓ data mismatch
		Variance	
Human error	0%	↑ Avoidable bias	↑ Avoidable bias
Training error	10%	↓ bias	10% ↓ variance
Training-dev error	11%		11% ↓ data mismatch
Dev error	12%		20% ↓ Bias + Data mismatch

Bias/variance on mismatched training and dev/test sets

Human level	4%		4%
Training set error	7%	↑ avoidable bias	7%
Training-dev set error	10%	↑ variance	10%
→ Dev error	12%	↑ data mismatch	6%
→ Test error	12%	↑ degree of overfitting to dev set.	6%

More general formulation

Reverse mirror

