



deeplearning.ai

Mismatched training and dev/test data

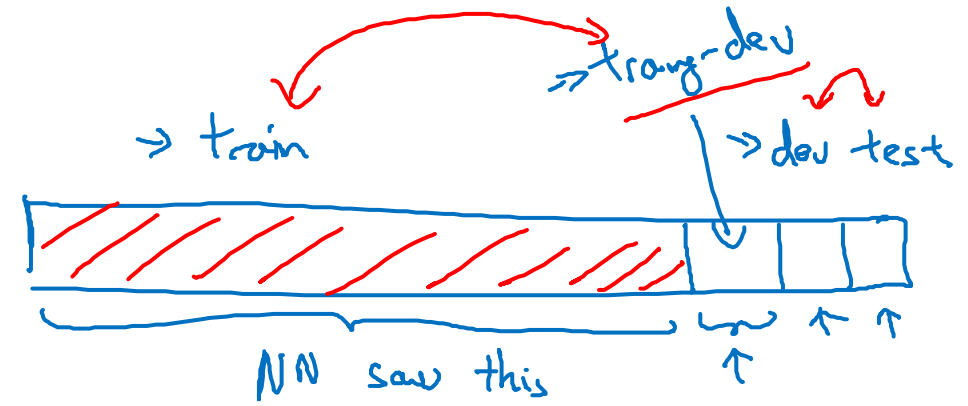
Bias and Variance with mismatched data distributions

Cat classifier example

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

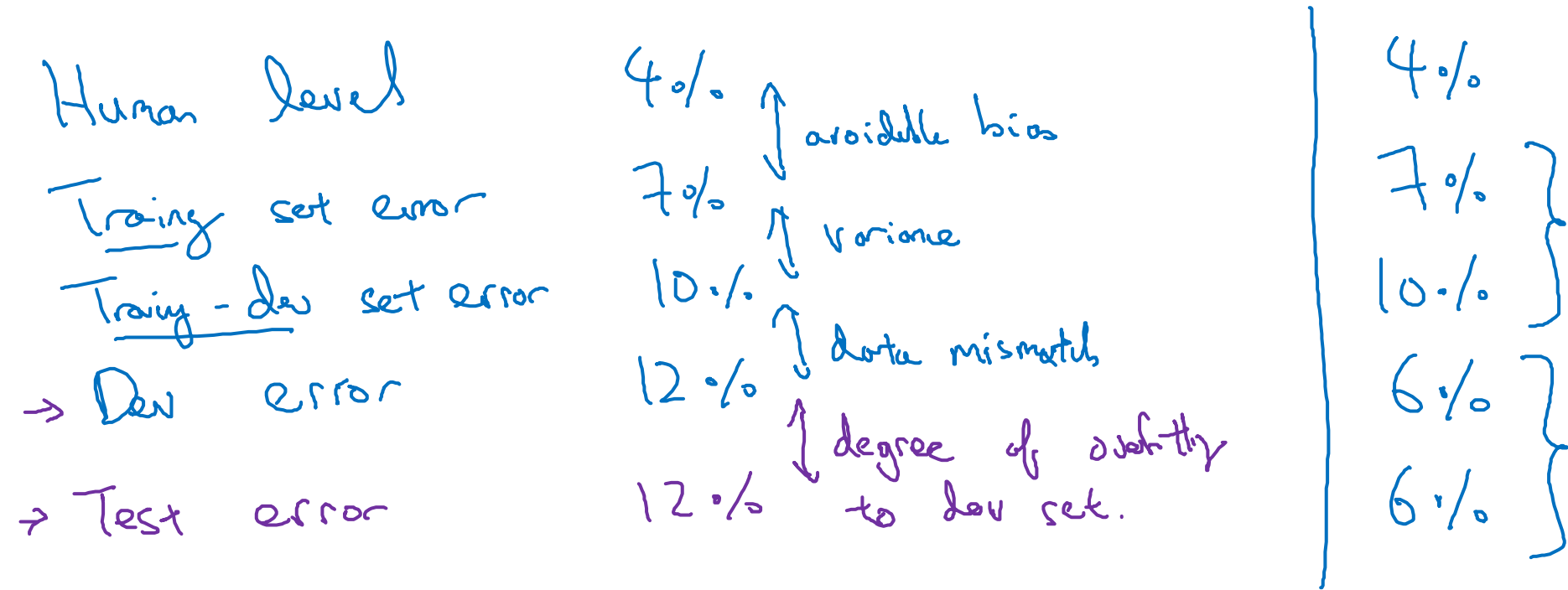
Training error 1%
 Dev error 10% $\downarrow 9\%$

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



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

Bias/variance on mismatched training and dev/test sets



More general formulation

Residual mirror

