



deeplearning.ai

# Mismatched training and dev/test data

---

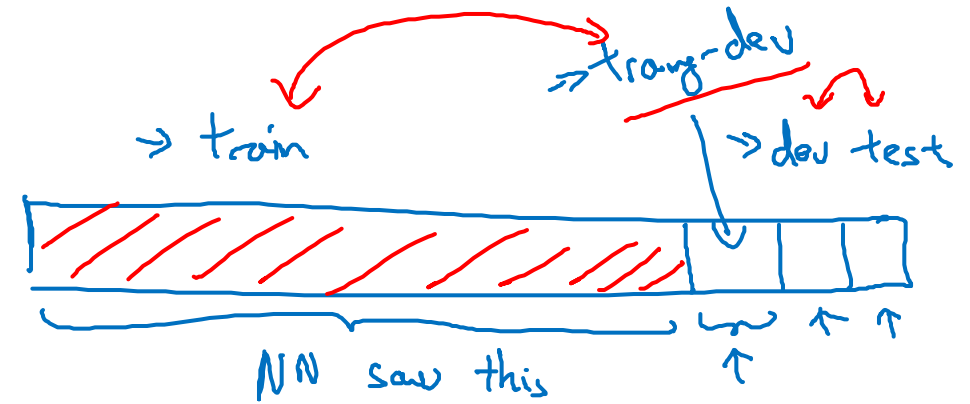
## Bias and Variance with mismatched data distributions

# Cat classifier example

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

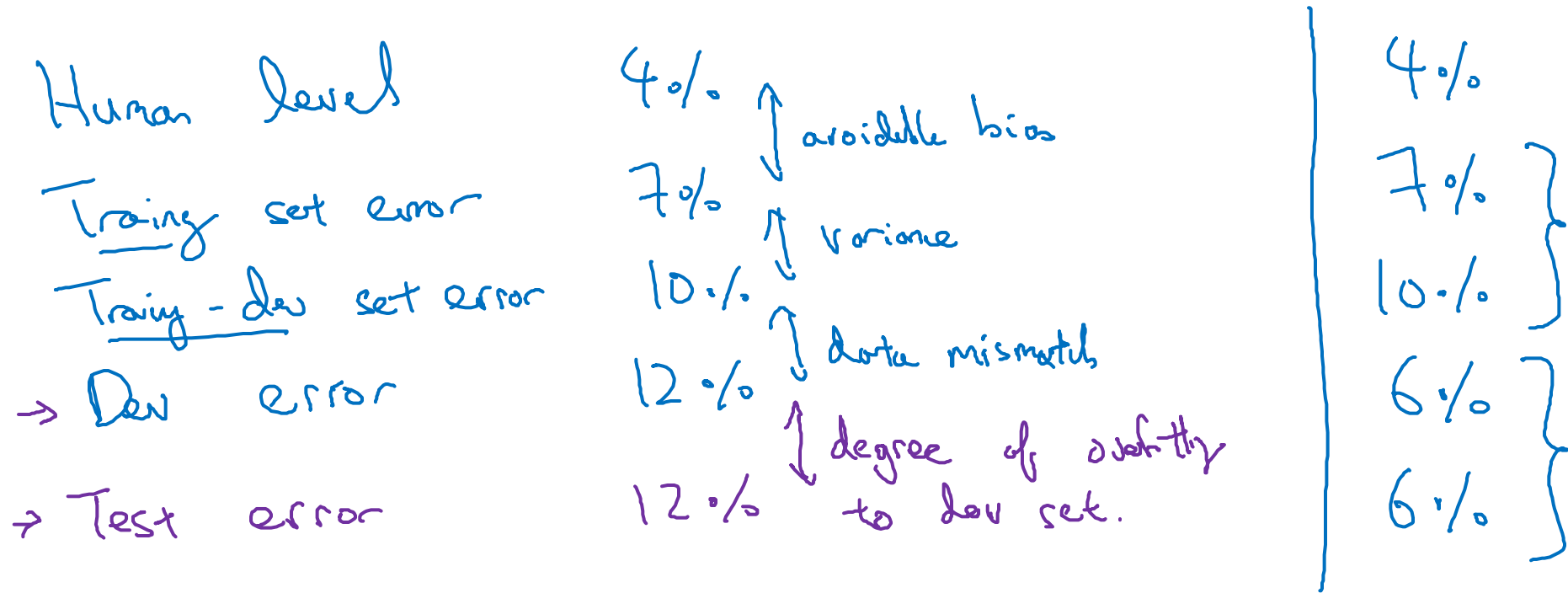
Training error .....  $1\%$   
 Dev error .....  $10\%$   $\downarrow 9\%$

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



Training error	$1\%$	$\uparrow$ Variance	$1\%$	$\uparrow$ Variance
→ Training-dev error	$9\%$		$1.5\%$	$\uparrow$ Data mismatch
→ Dev error	$10\%$		$10\%$	
		Variance		
Human error	..... $0\%$	$\uparrow$ Avoidable bias		$\uparrow$ Avoidable bias
Training error	$10\%$	$\downarrow$ bias	$10\%$	$\downarrow$ Variance
Training-dev error	$11\%$		$11\%$	$\uparrow$ Data mismatch
Dev error	$12\%$		$20\%$	
	Bias		Bias + Data mismatch	

# Bias/variance on mismatched training and dev/test sets



# More general formulation

Recurrent mirror

