DATA

TRAIN	DEV	TEST

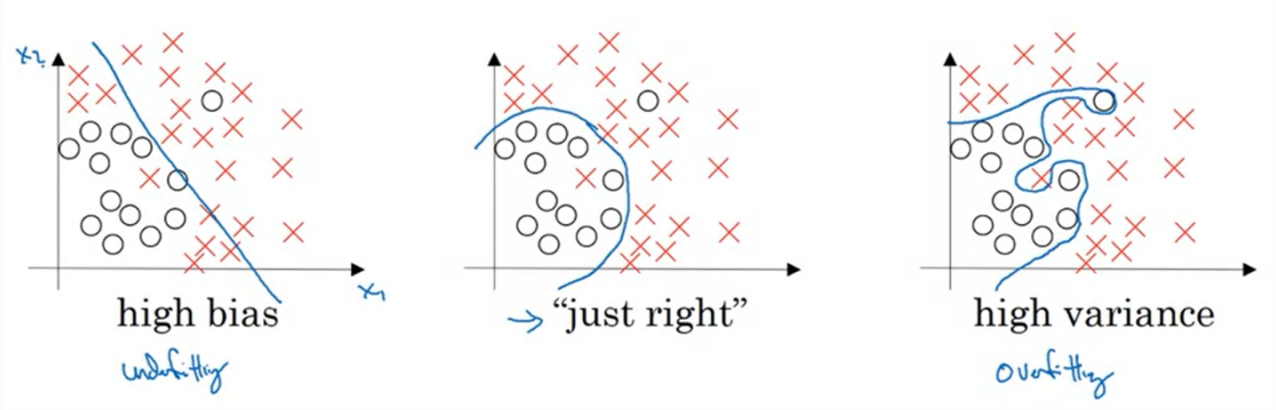
Dev: test multiple algorithms on it, select best.

Test: check for that chosen algorithm.

In big data: 99% - 0'5% - 0'5%

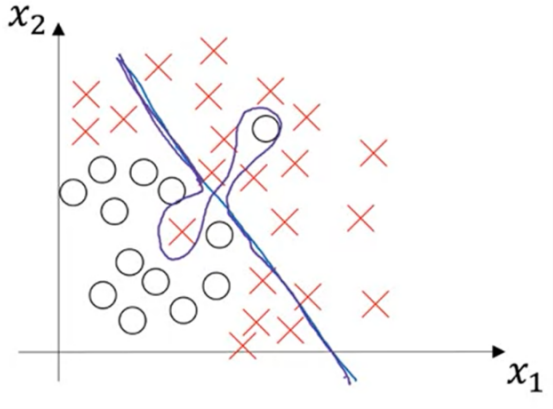
AVOID mismatched data samples!

BIAS/VARIANCE



Train set error	1%	15%	15%	0'5%
Dev set error	11%	16%	30%	1%
	high variance	high bias	high bias high variance	low bias low variance

High bias and high variance



- Do I have high bias? Check for training data performance.
  - Bigger network
  - Train larger
- Do I have high variance? Check for dev set performance.
  - More data
  - Regularization

Observations:

- The model with He initialization separates the blue and the red dots very well in a small number of iterations.

---

7 - Conclusions

You've tried three different types of initializations. For the same number of iterations and same hyperparameters, the comparison is:

Model	Train accuracy	Problem/Comment
3-layer NN with zeros initialization	50%	fails to break symmetry
3-layer NN with large random initialization	83%	too large weights
3-layer NN with He initialization	99%	recommended method

Congratulations! You've completed this notebook on Initialization.

Here's a quick recap of the main takeaways:

- Different initializations lead to very different results
- Random initialization is used to break symmetry and make sure different hidden units can learn different things
- Resist initializing to values that are too large!
- He initialization works well for networks with ReLU activations