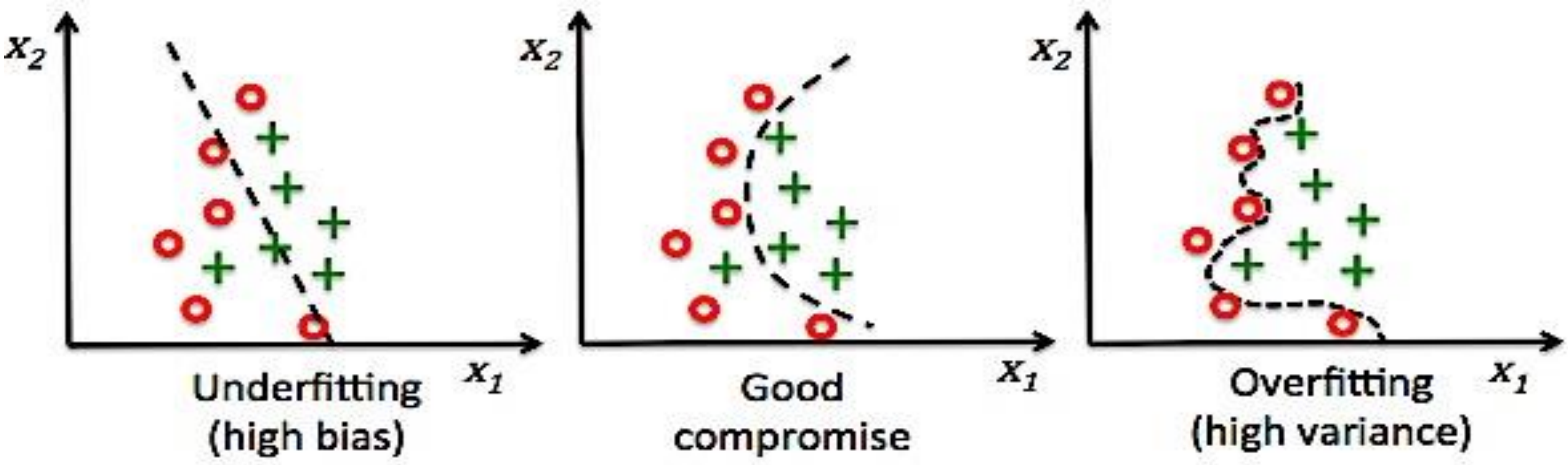
Underfitting and Overfitting

Bias/Variance



Bias/Variance

The algorithm's error rate on the training set is algorithm's **bias** .

How much worse the algorithm does on the dev (or test) set than the training set is algorithm's **variance** .

Bias/Variance

Train Set Error	1	12	8	1
Dev Set Error	9	13	16	1.5
	High Variance	High Bias	High Bias, High Variance	Low Bias Low Variance
	Overfitting	Underfitting	Underfitting	Good fit

Multi dimensional system can have high bias in some areas and high variance in some other areas of the system, resulting in High Bias and High Variance issue

High Bias

Increase the model size (such as number of neurons/layers)	It allows to fit the training set better. If you find that this increases variance, then use regularization, which will usually eliminate the increase in variance.
Modify input features based on insights from error analysis	Create additional features that help the algorithm eliminate a particular category of errors. These new features could help with both bias and variance.
Reduce or eliminate regularization (L2, L1 regularization, dropout)	reduces avoidable bias, but increase variance.
Modify model architecture (such as neural network architecture) so that it is more suitable for your problem	This can affect both bias and variance.

High Variance

Add more training data	Simplest and most reliable way to address variance, so long as you have access to significantly more data and enough computational power to process the data.
Add regularization (L2, L1 regularization, dropout)	This technique reduces variance but increases bias.
Add early stopping (stop gradient descent early, based on dev set error)	Reduces variance but increases bias.
Modify model architecture (such as neural network architecture) so that it is more suitable for your problem	This affects both bias and variance.

Often compare with human level performance

Image recognition, spam classification.	Ease of obtaining data from human labelers
	Error analysis can draw on human intuition.
	Use human-level performance to estimate the optimal error rate and also set a “desired error rate

Tasks Where we don't compare with human level performance

Picking a book to recommend to you; picking an ad to show a user on a website; predicting stock market.	It is harder to obtain labels
	Human intuition is harder to count on
	It is hard to know what the optimal error rate and reasonable desired error rate is

Classification example for animals

Type	Scenario 1	Scenario 2	
Humans (Bayes error)	1	7.5	
Training error	8	8	
Dev error	10	10	
	Focus on Bias	Focus on Variance	
Avoidable Bias	7	0.5	

New scenario

Type	Scenario 1	Scenerio2	
Human (Bayes) error	1	1	1
	0.7	0.7	0.7
	0.5	0.5	0.5
Training error	5	1	0.7
Dev error	6	5	0.8
	Bias issue	Variance Issue	Difficult

Two fundamental Assumptions:	You can fit the training set well
	Training set performance should Generalize to dev/test set.