Bias and variance

And how they are related to under/over fitting

Sijin Zhang

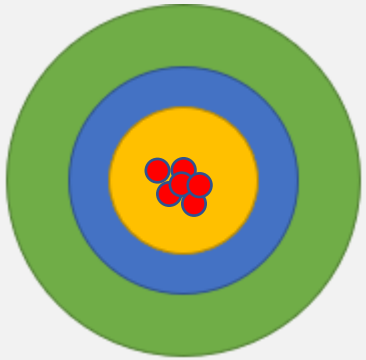
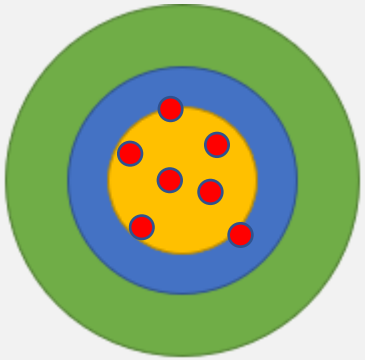
What are bias and variance

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it’s error (e.g., squared error)

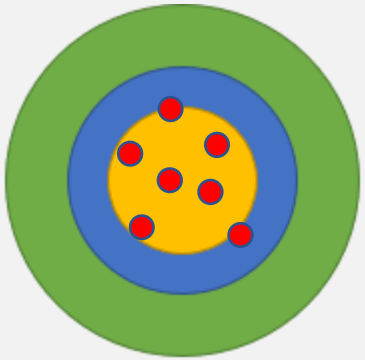
What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

		Variance	
		Low	High
Bias	Low		
	High		

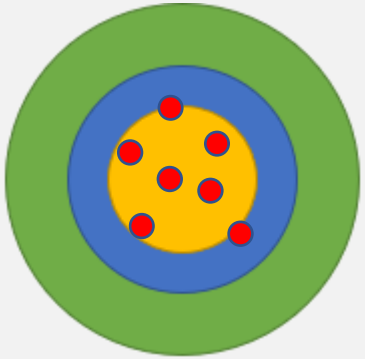
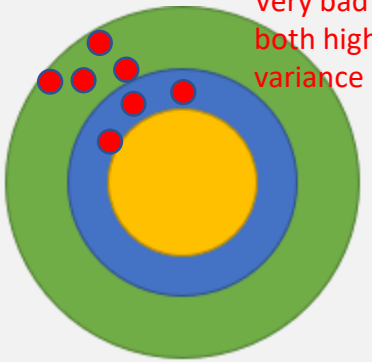
What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

		Variance	
		Low	High
Bias	Low	Usually means that we are making a pretty good prediction 	
	High		

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

		Variance	
		Low	High
Bias	Low	Usually means that we are making a pretty good prediction 	
	High		Very bad case, if we have both high bias and high variance 

What are bias and variance

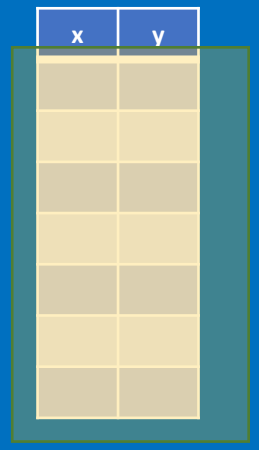
Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

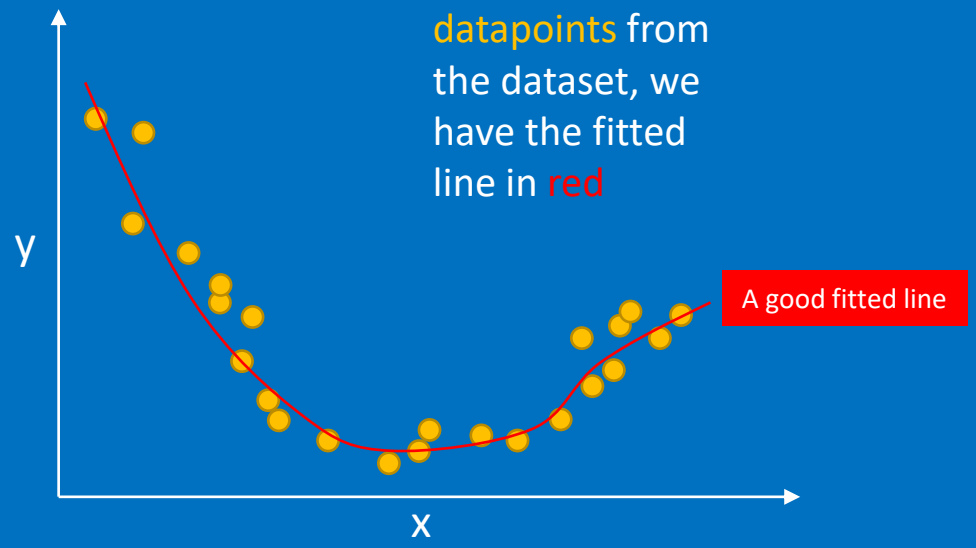
Assuming we have a dataset, with one predictor, and one target

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)



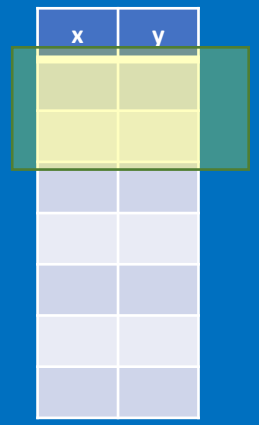
Assuming we have a dataset, with one predictor, and one target



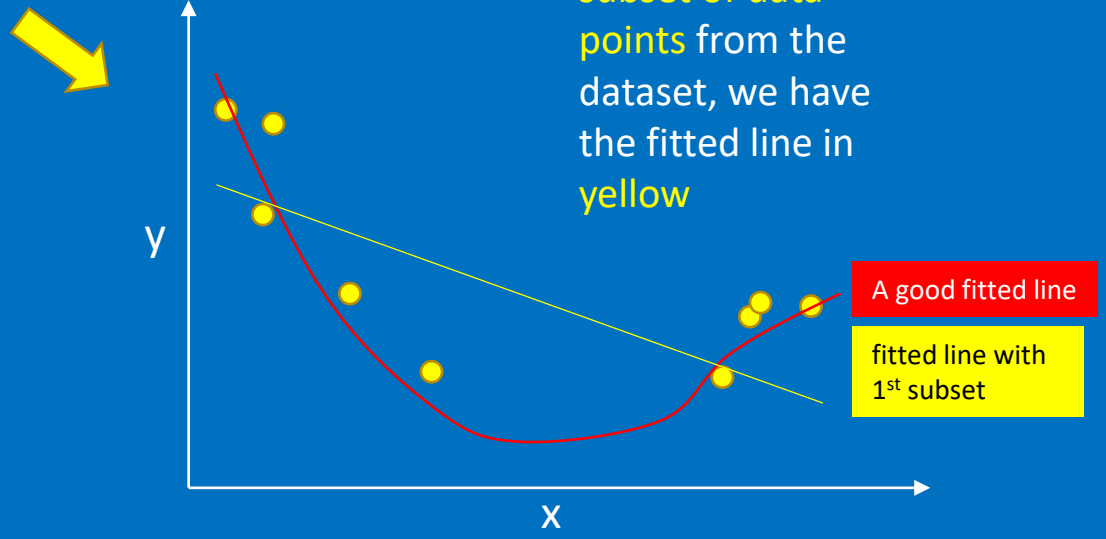
If we take all datapoints from the dataset, we have the fitted line in red

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)



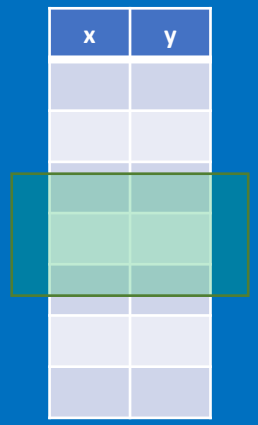
Assuming we have a dataset, with one predictor, and one target



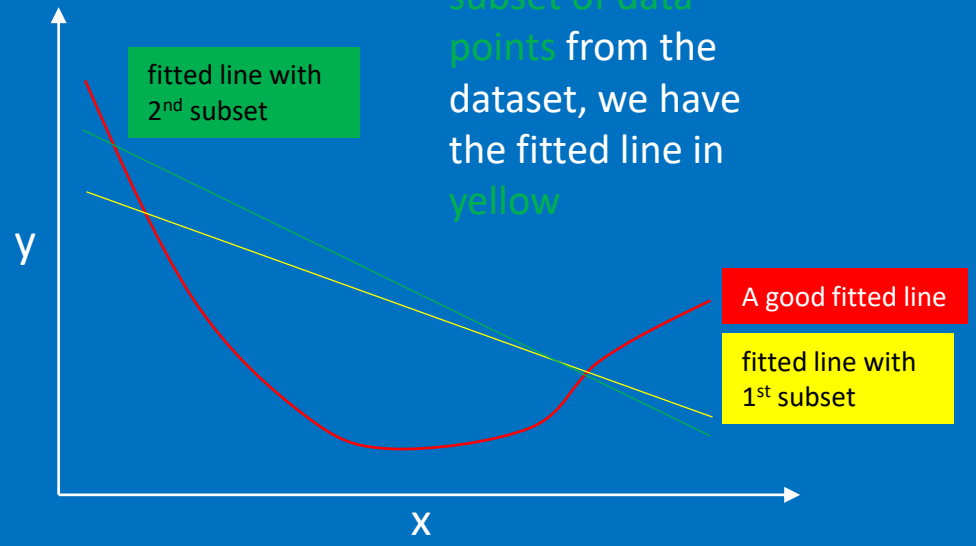
If we take a subset of data points from the dataset, we have the fitted line in yellow

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)



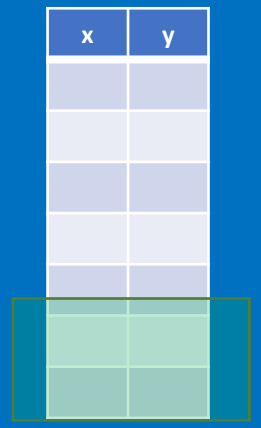
Assuming we have a dataset, with one predictor, and one target



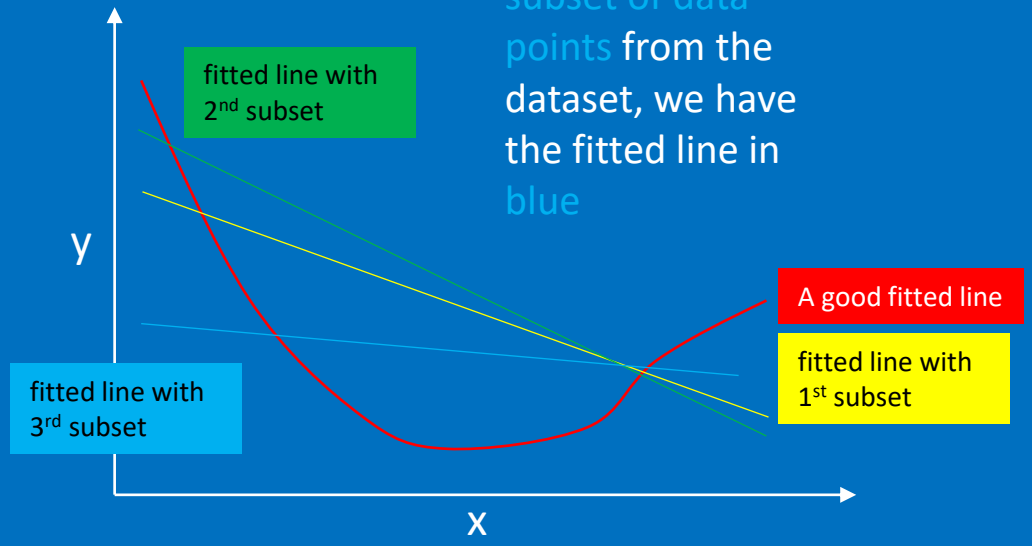
If we take another subset of data points from the dataset, we have the fitted line in yellow

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)



Assuming we have a dataset, with one predictor, and one target



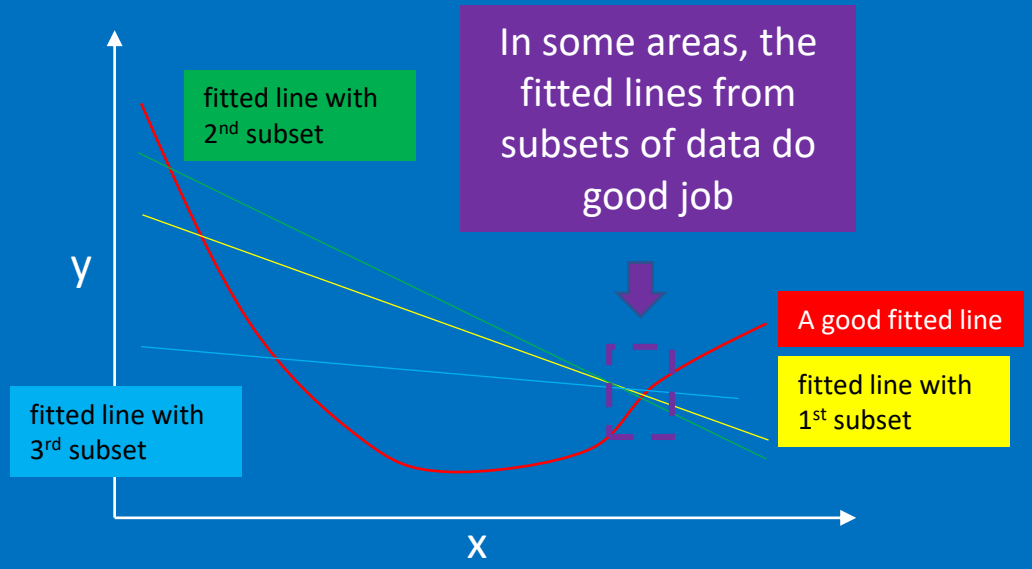
If we take the last subset of data points from the dataset, we have the fitted line in blue

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

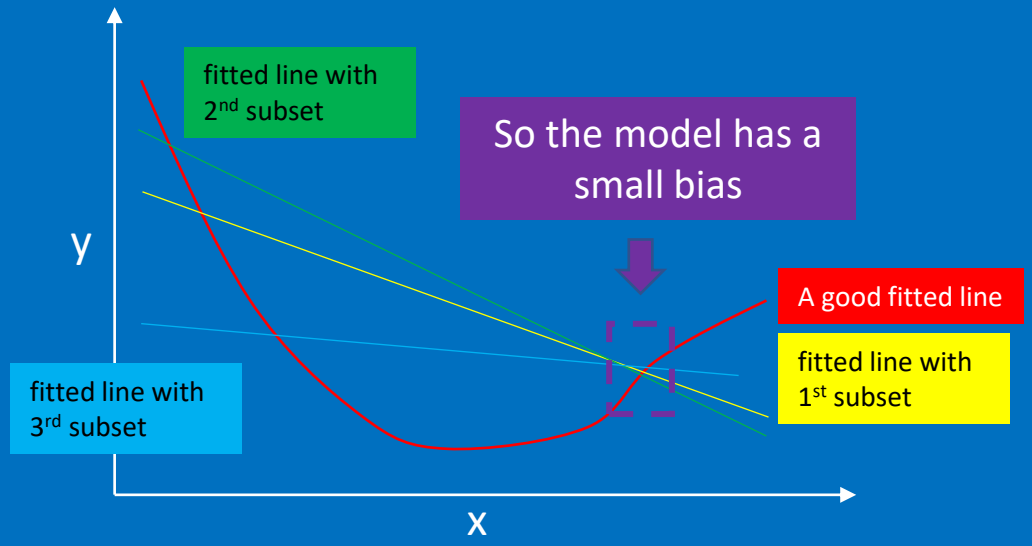


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

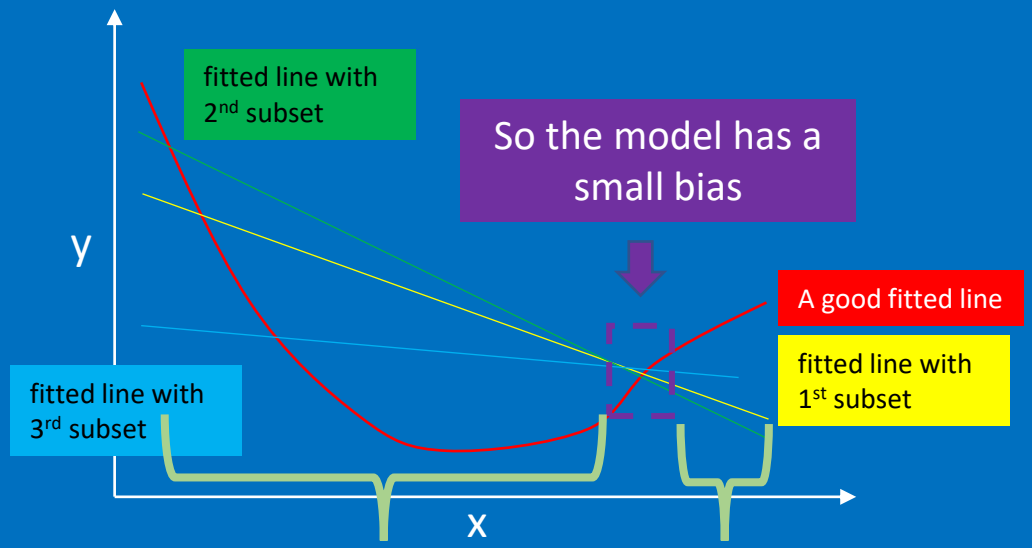


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

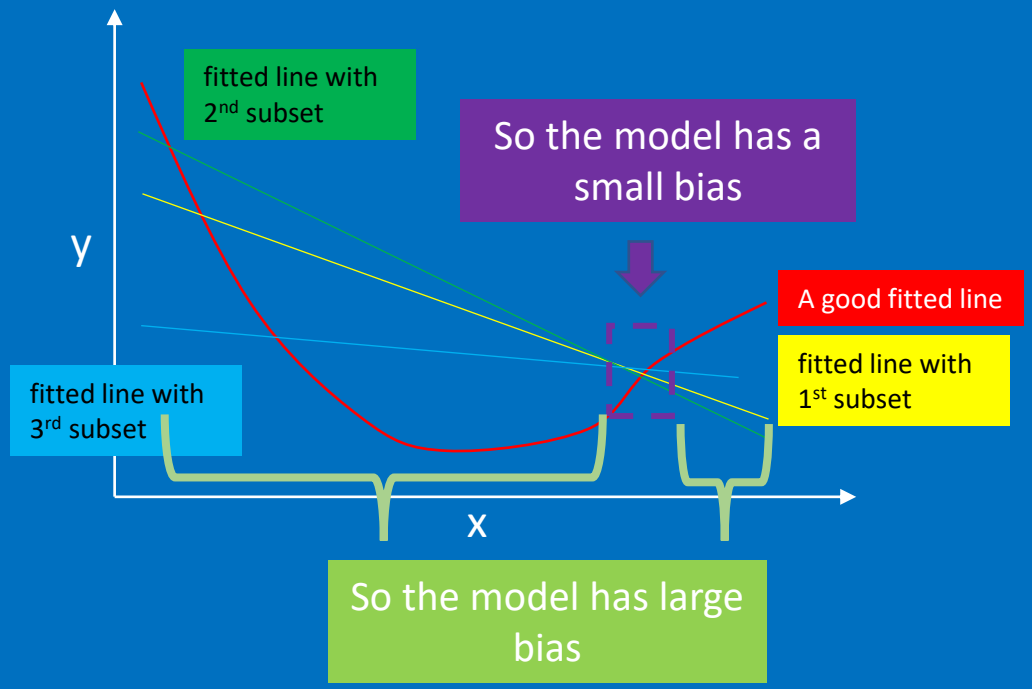


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

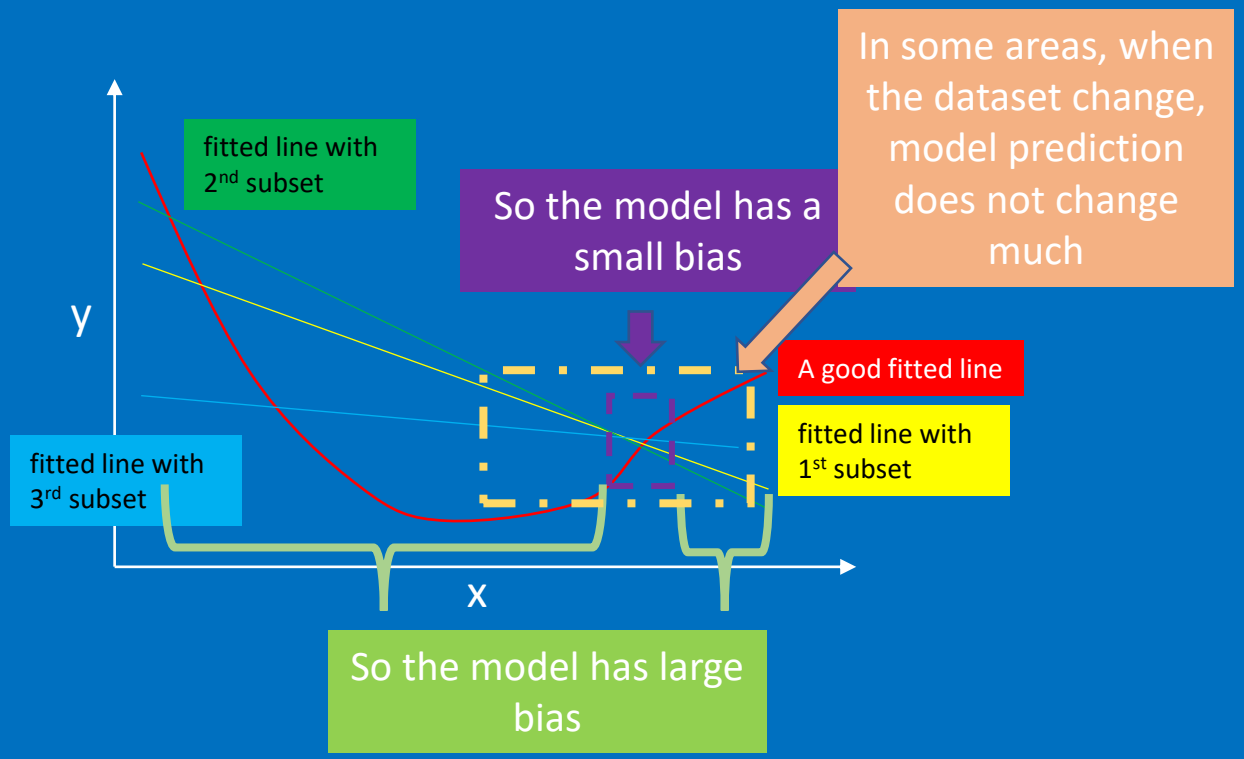


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

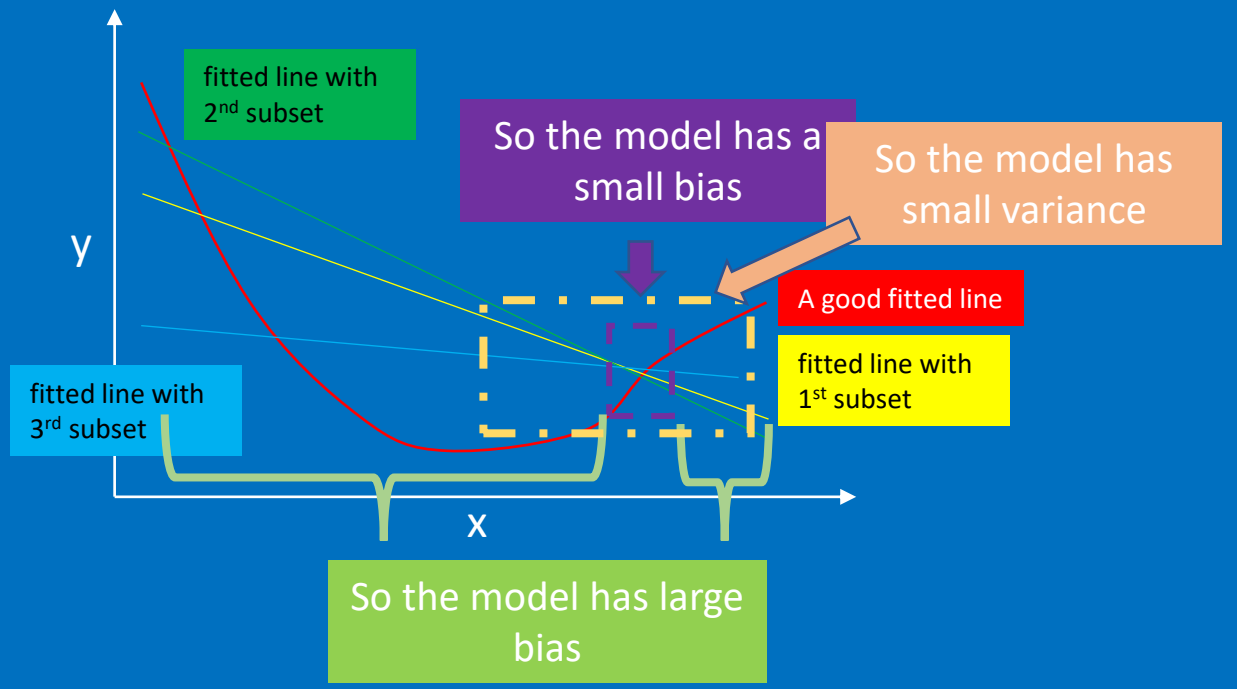


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

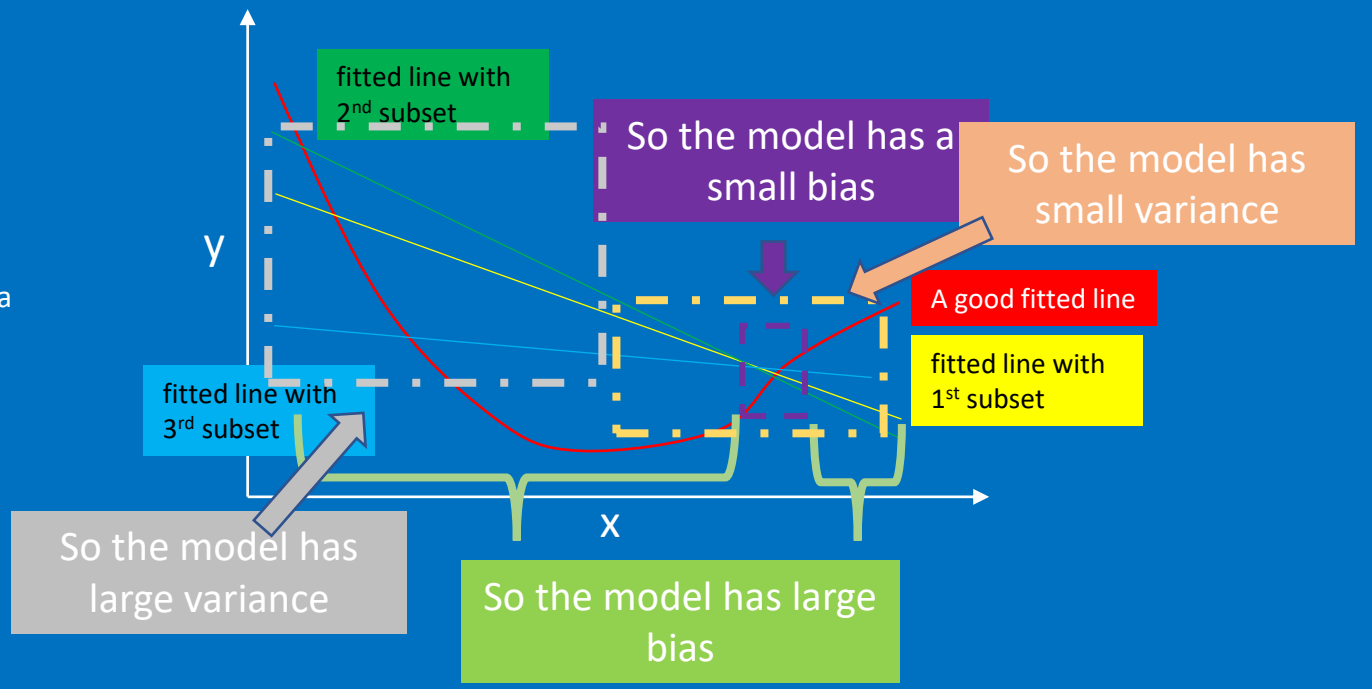


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

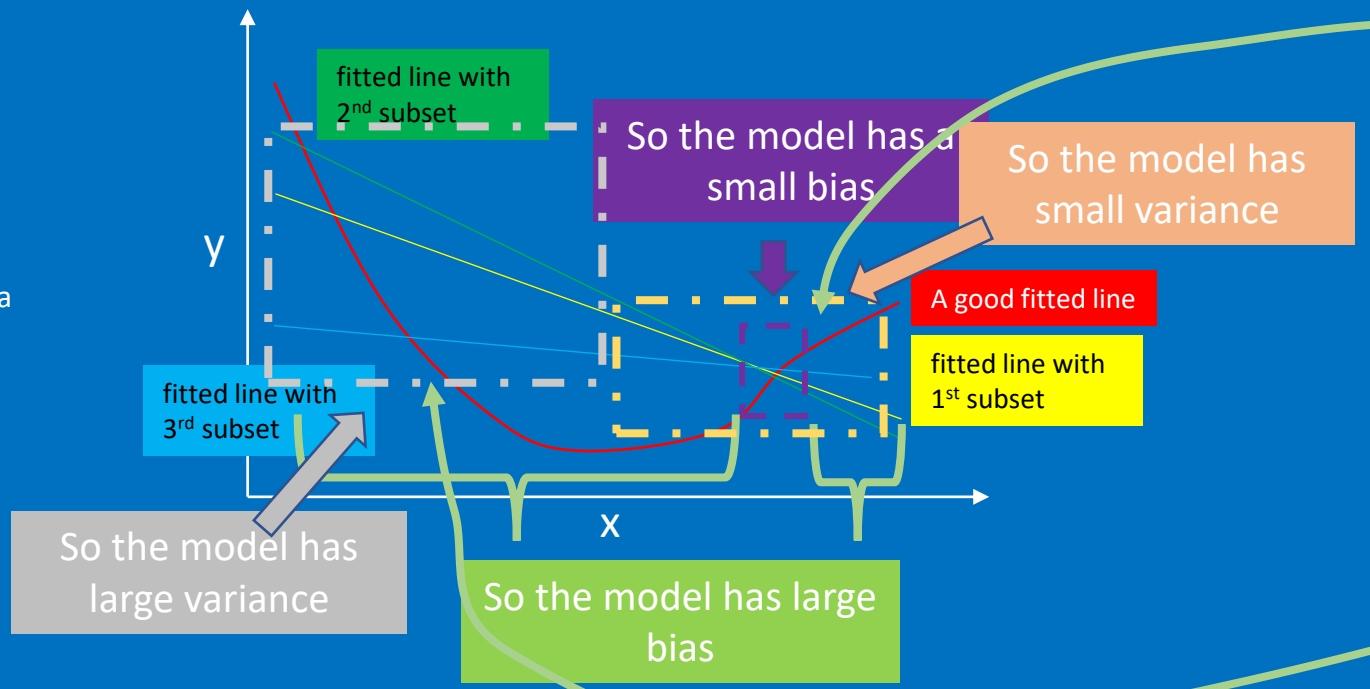


What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target



In a word,

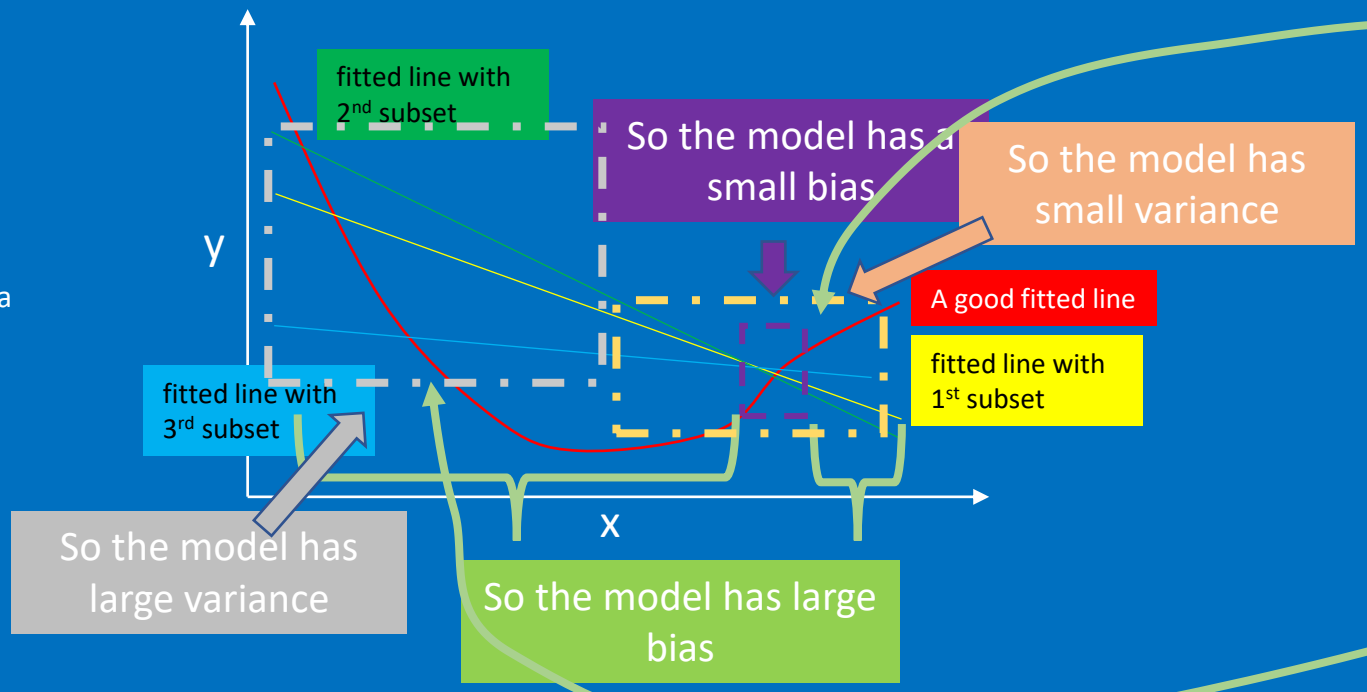
- the model is great if it has:
  Small bias + small variance
- The model is horrible if it has:
  Large bias + large variance

What are bias and variance

Bias and variance are two ways to describe the machine learning model and it's error (e.g., squared error)

x	y

Assuming we have a dataset, with one predictor, and one target

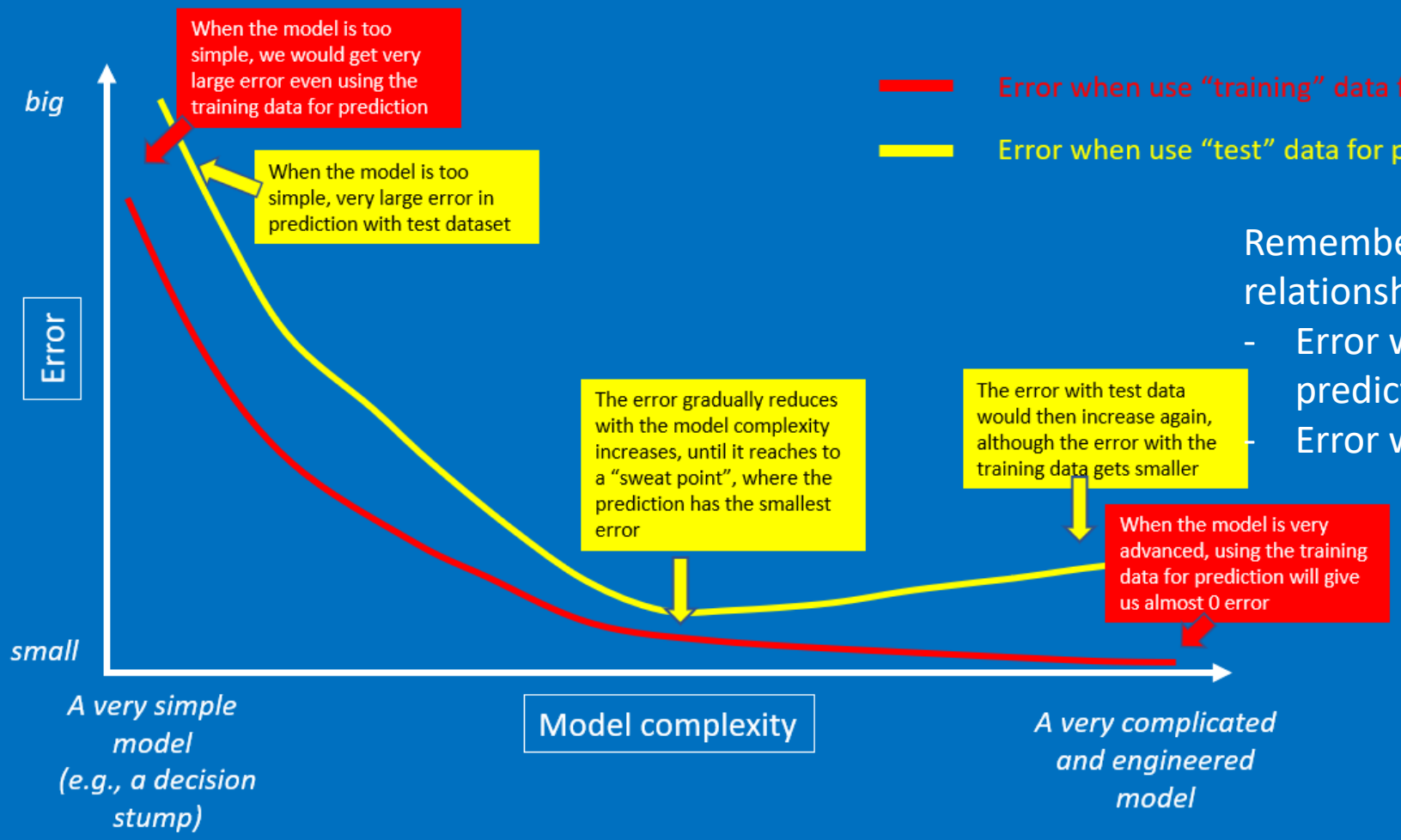


In a word,

- the model is great if it has:
  Small bias + small variance
- The model is horrible if it has:
  Large bias + large variance

Any ensemble method (e.g., bagging) would be a great way to reduce variance

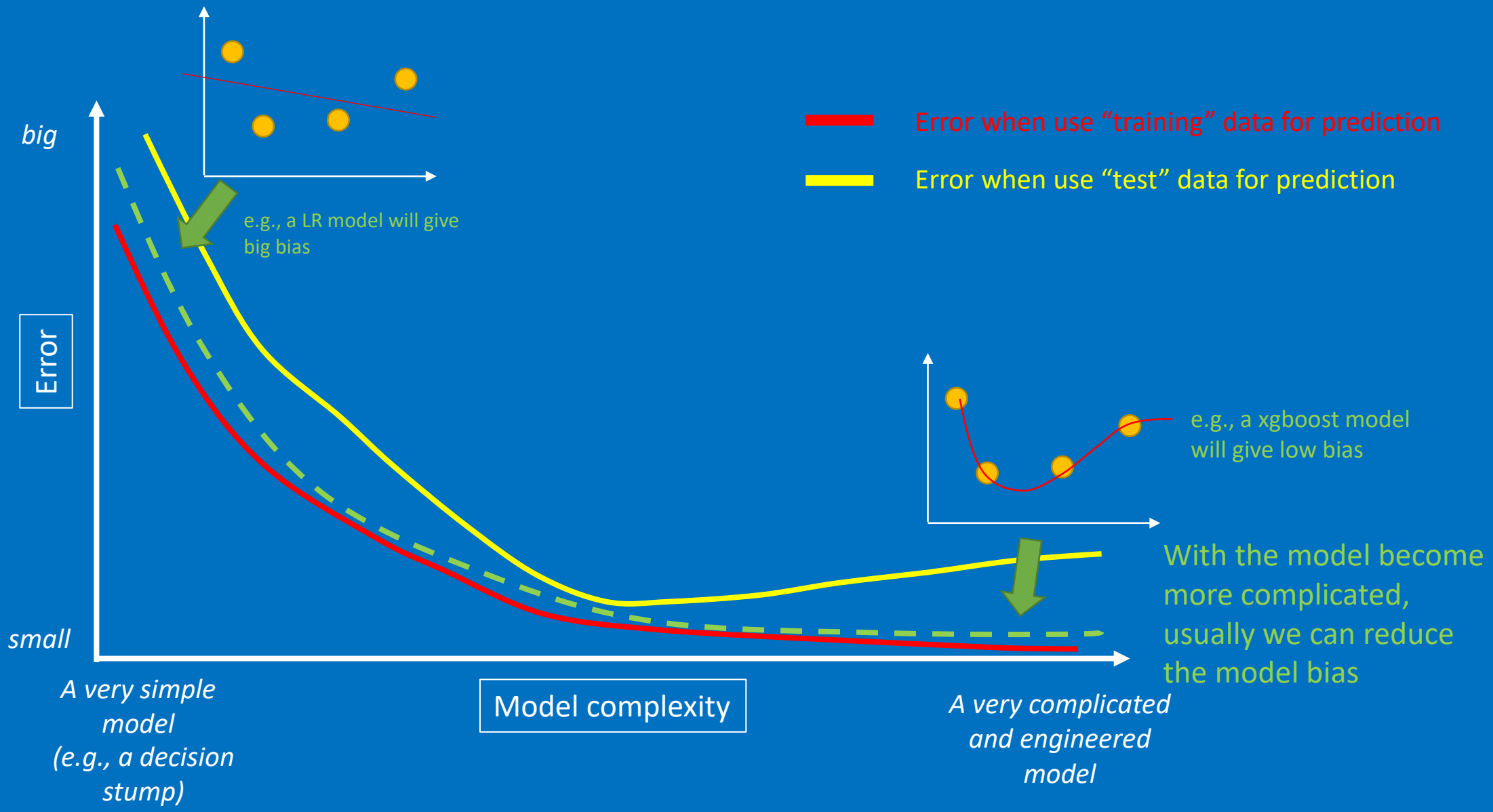
How bias and variance are related to under and overfitting



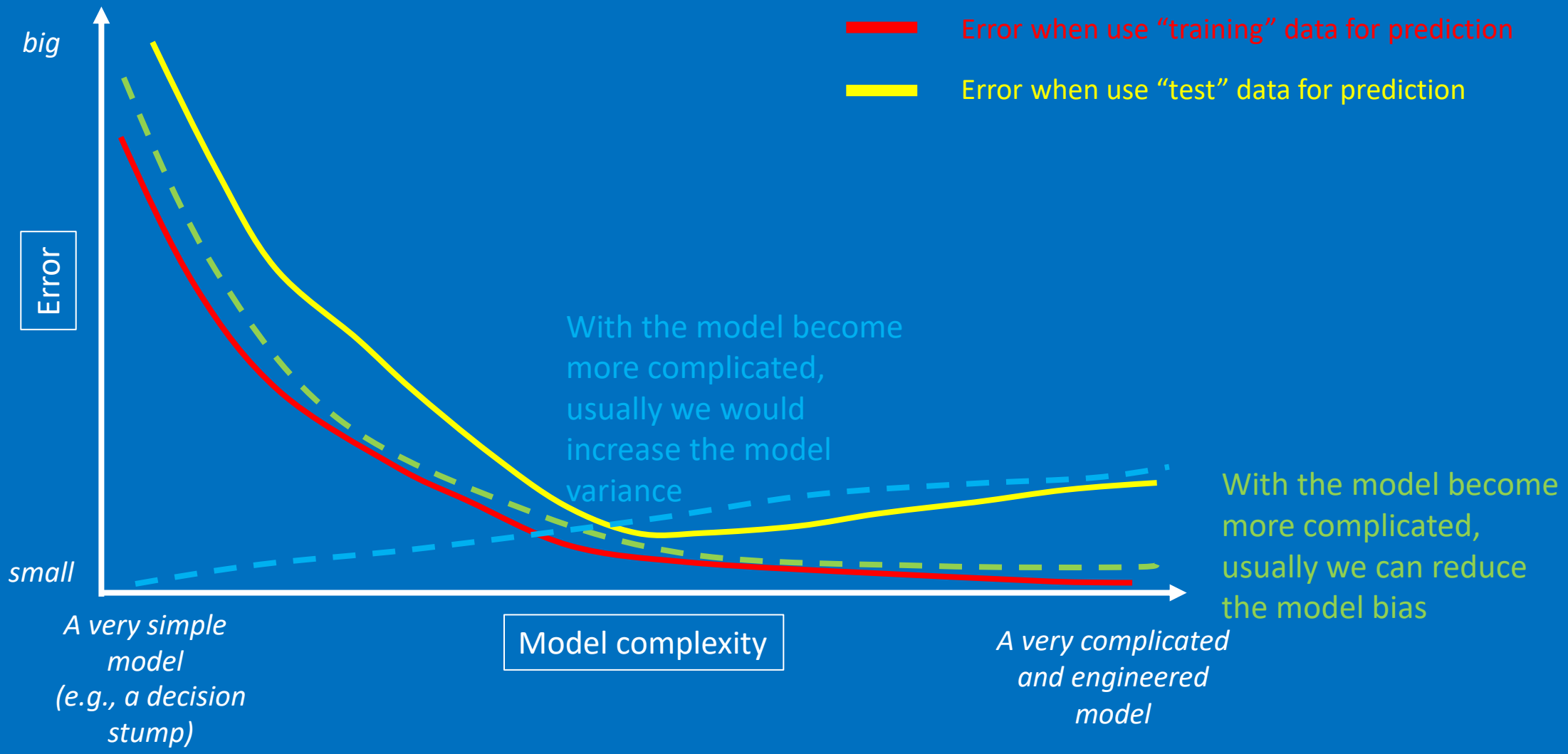
Remember this graph, which represents the relationships between model complexity and:

- Error when using "training" data for prediction
- Error when using "test" data for prediction

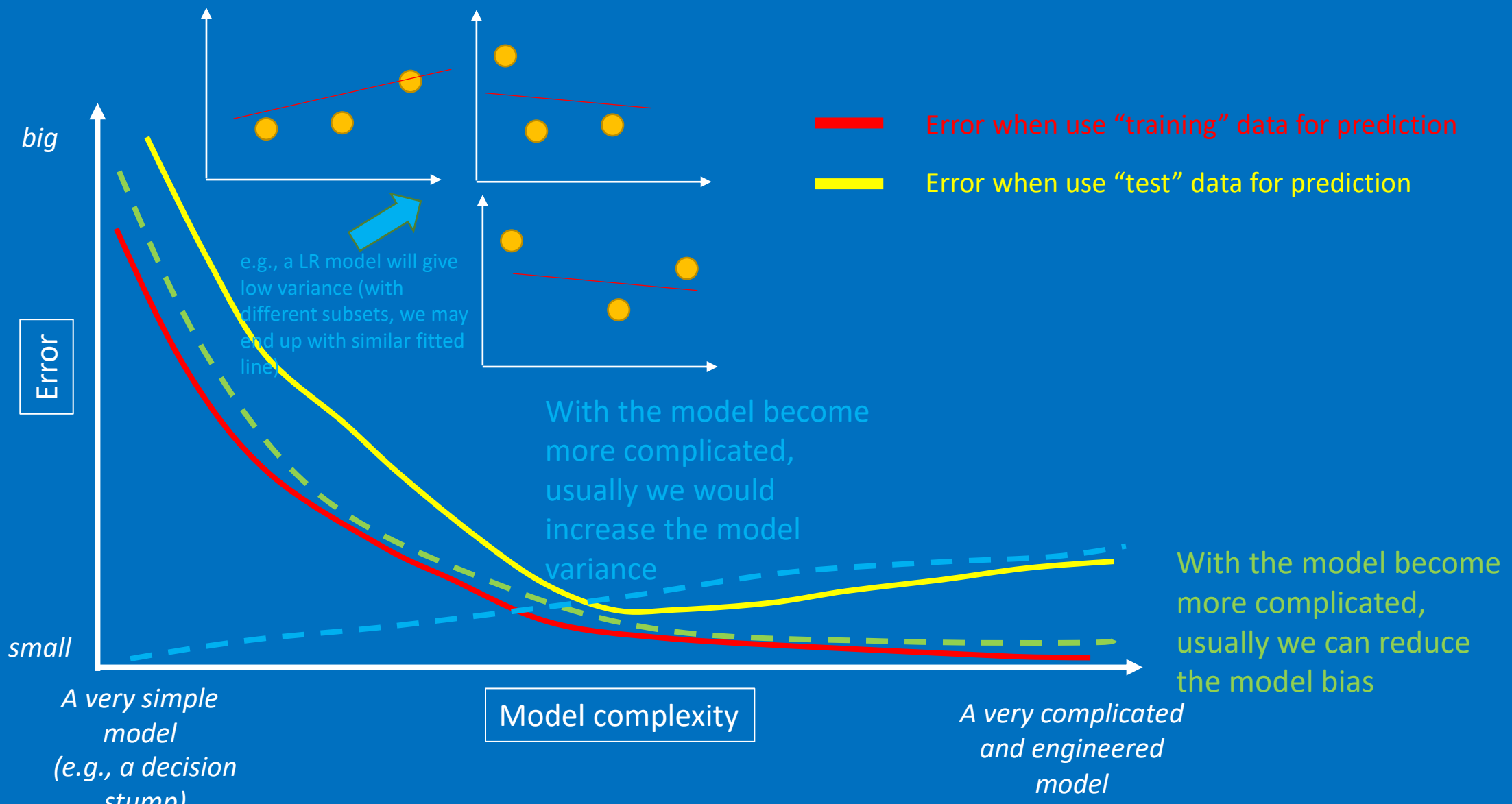
How bias and variance are related to under and overfitting



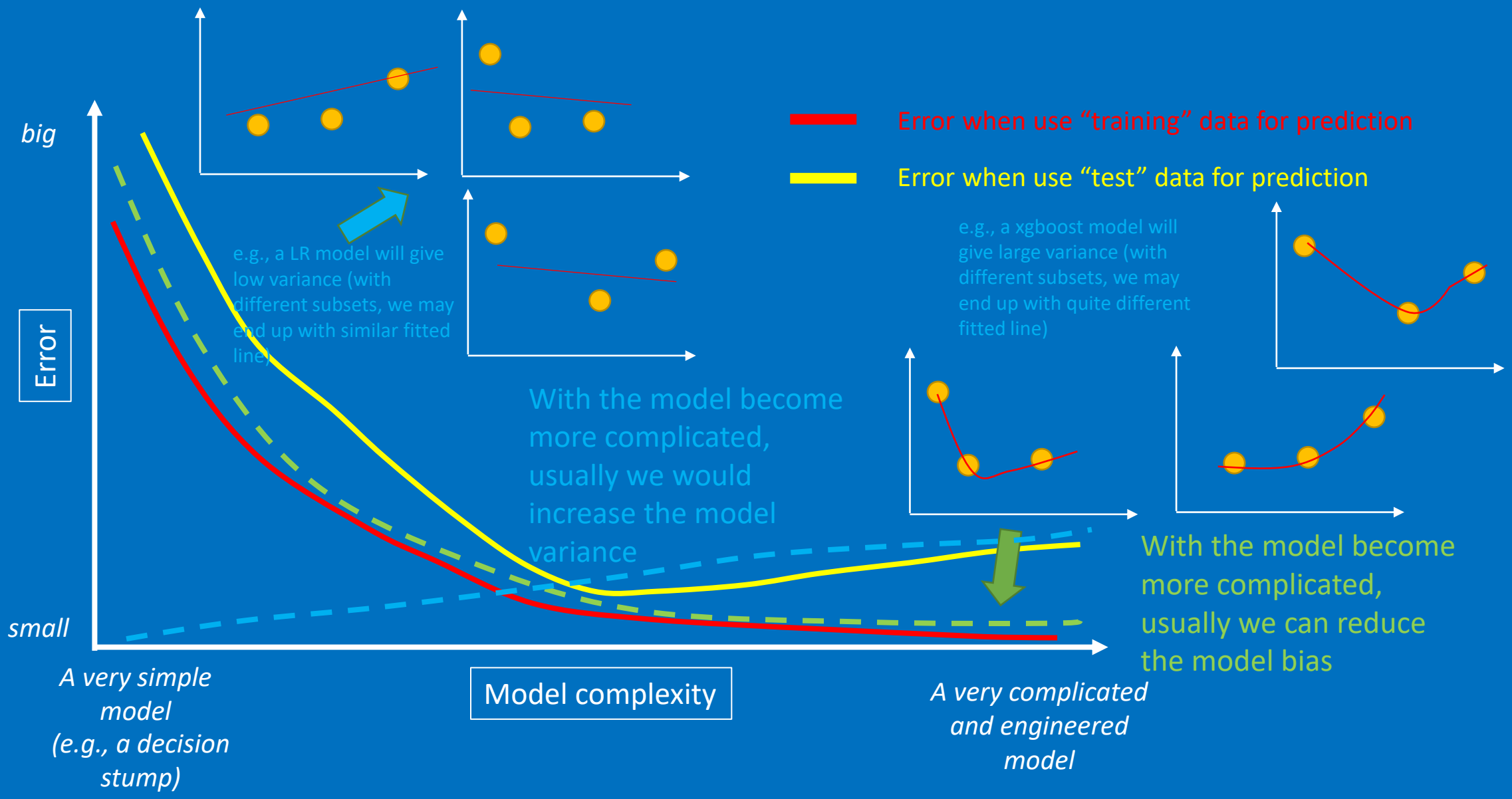
How bias and variance are related to under and overfitting



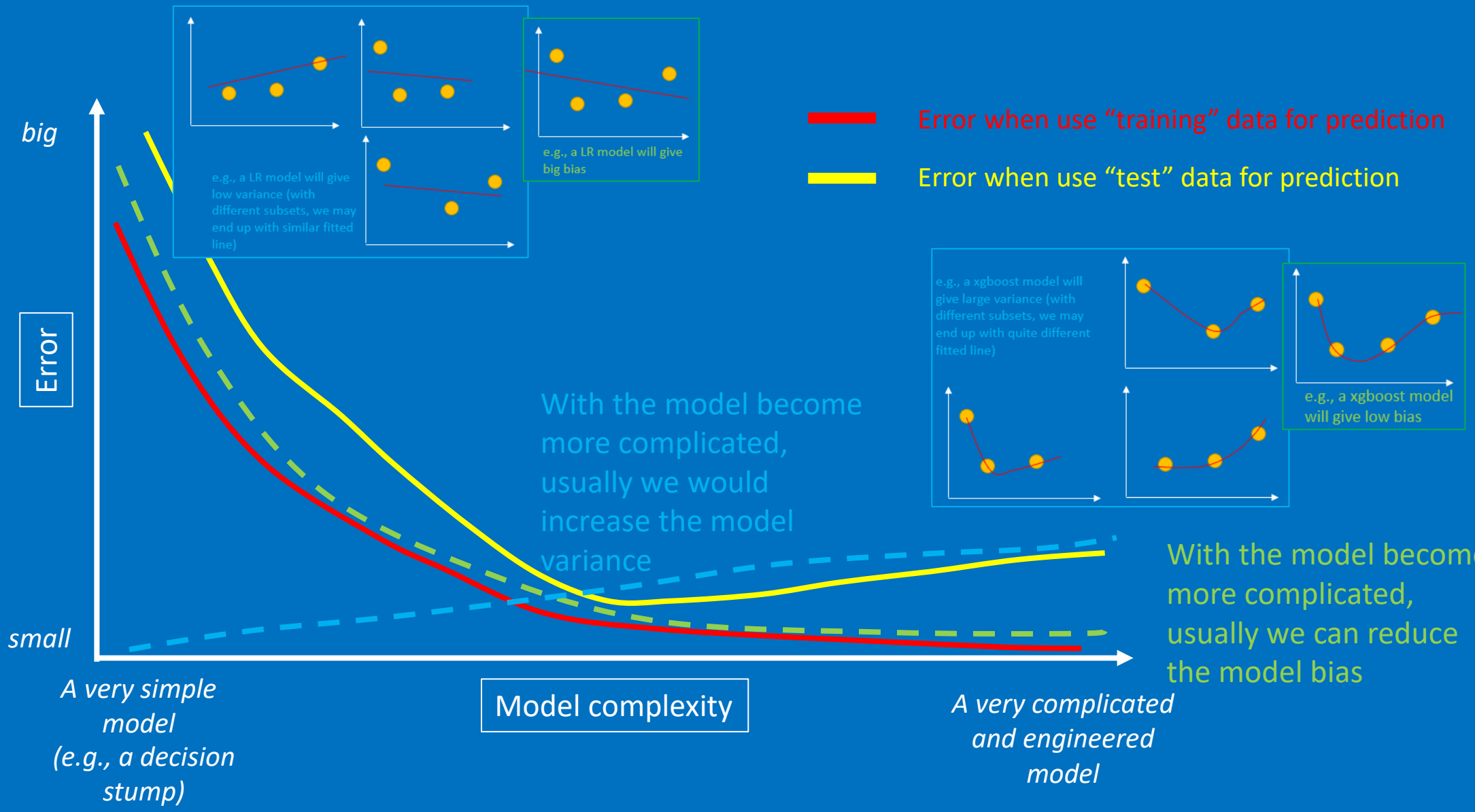
How bias and variance are related to under and overfitting



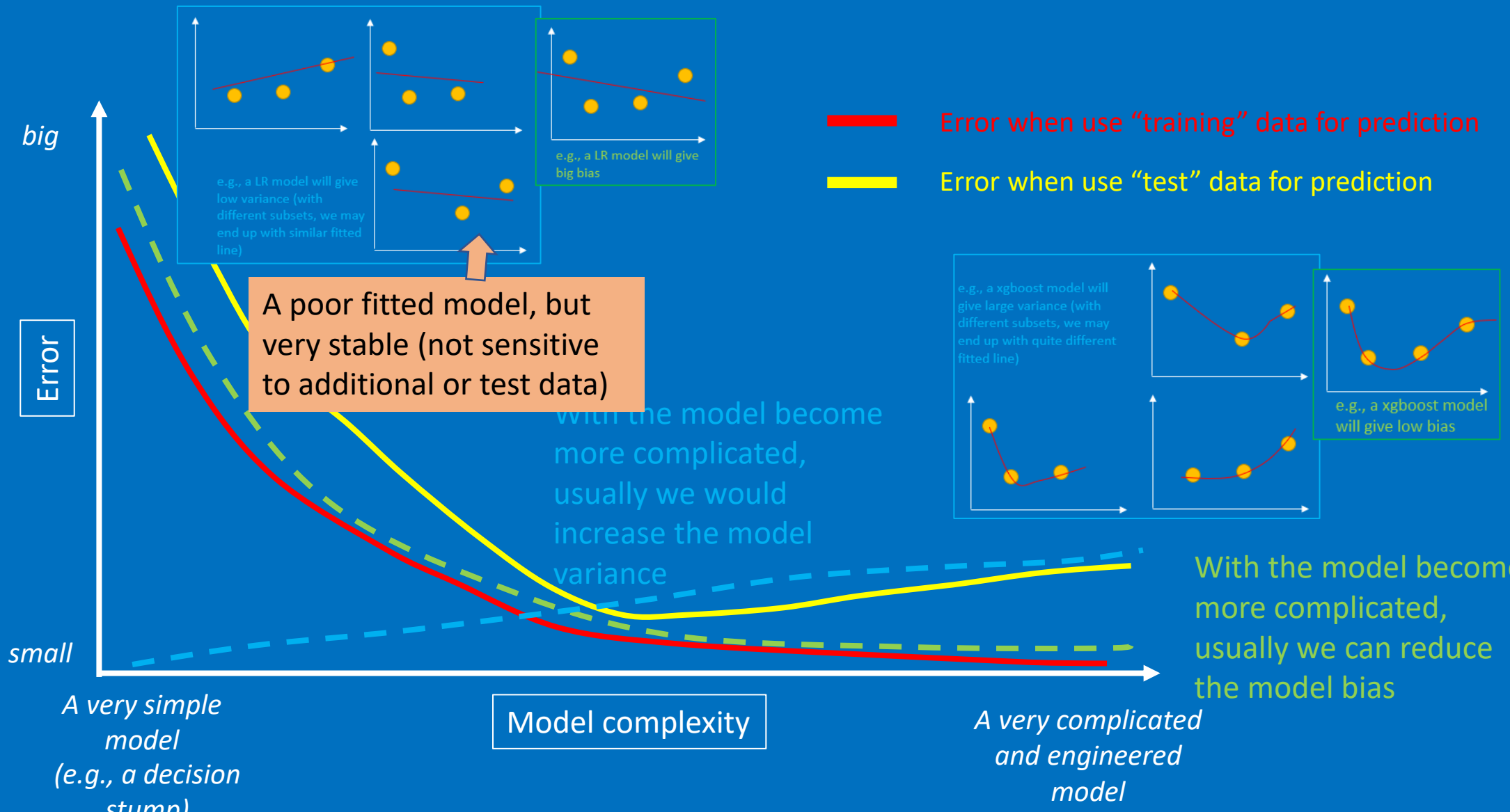
How bias and variance are related to under and overfitting



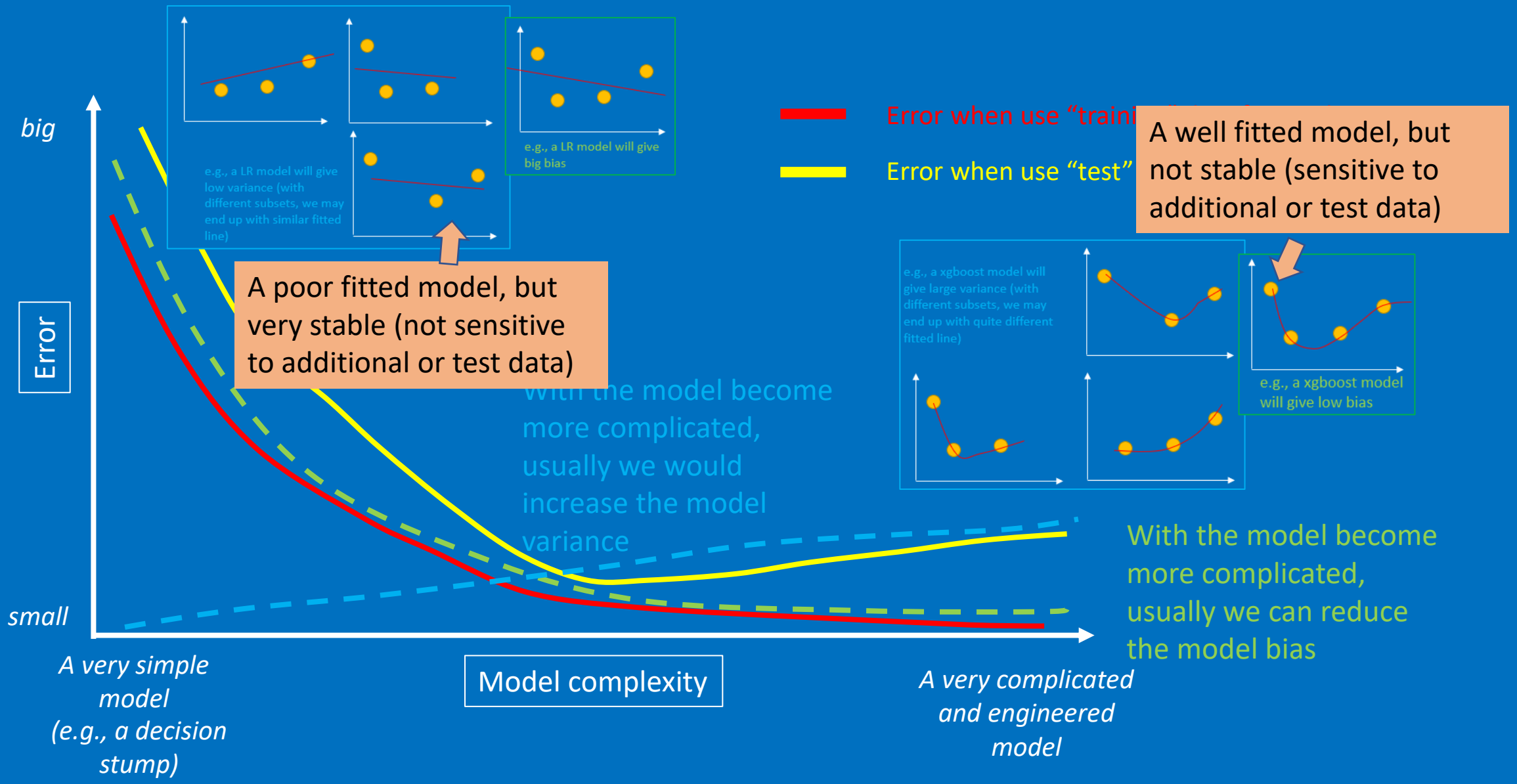
How bias and variance are related to under and overfitting



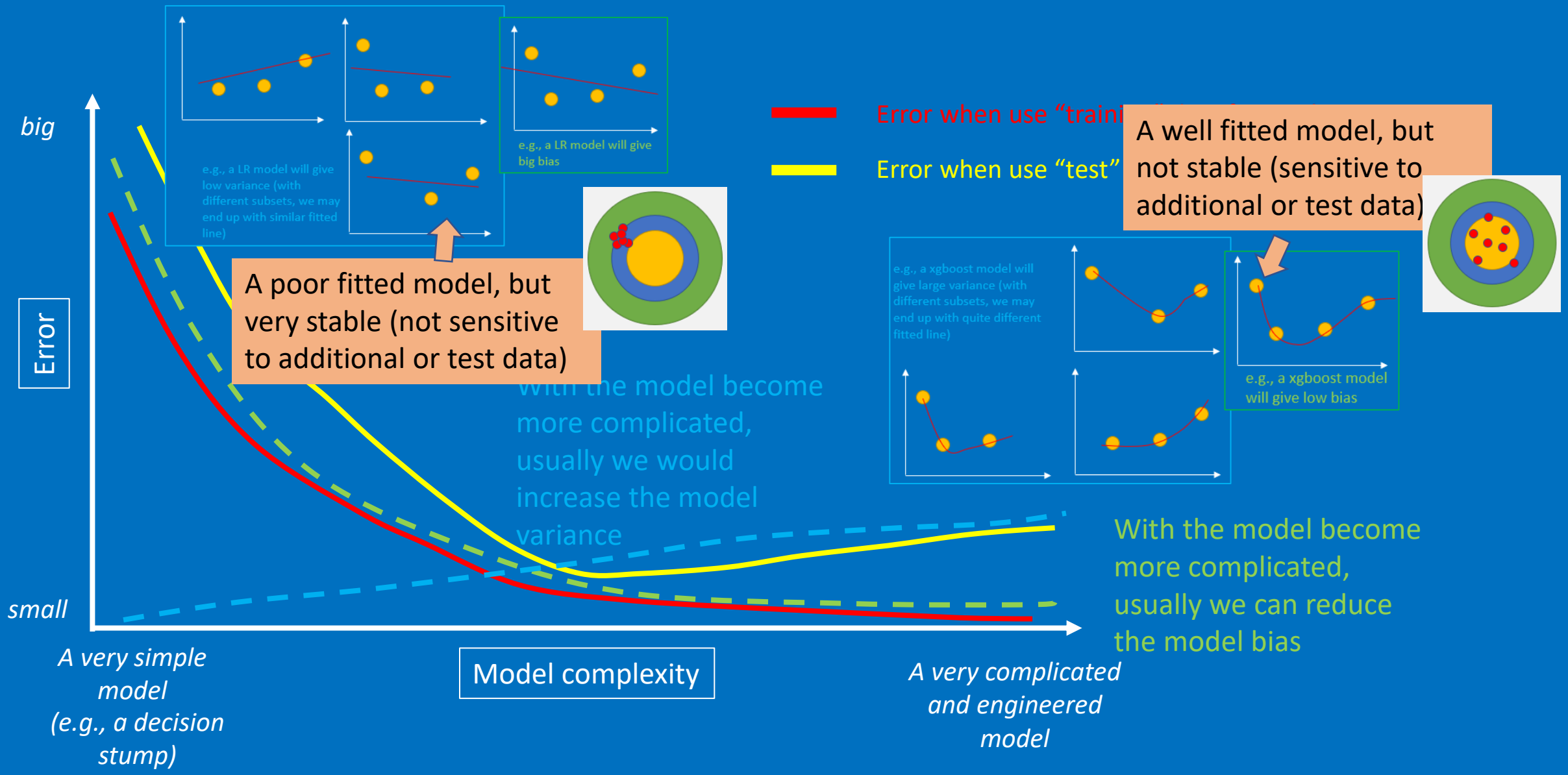
How bias and variance are related to under and overfitting



How bias and variance are related to under and overfitting



How bias and variance are related to under and overfitting



How bias and variance are related to under and overfitting