Generalization Error and Bias Variance Tradeoff

CS109A Introduction to Data Science

Pavlos Protopapas, Natesh Pillai and Chris Gumb

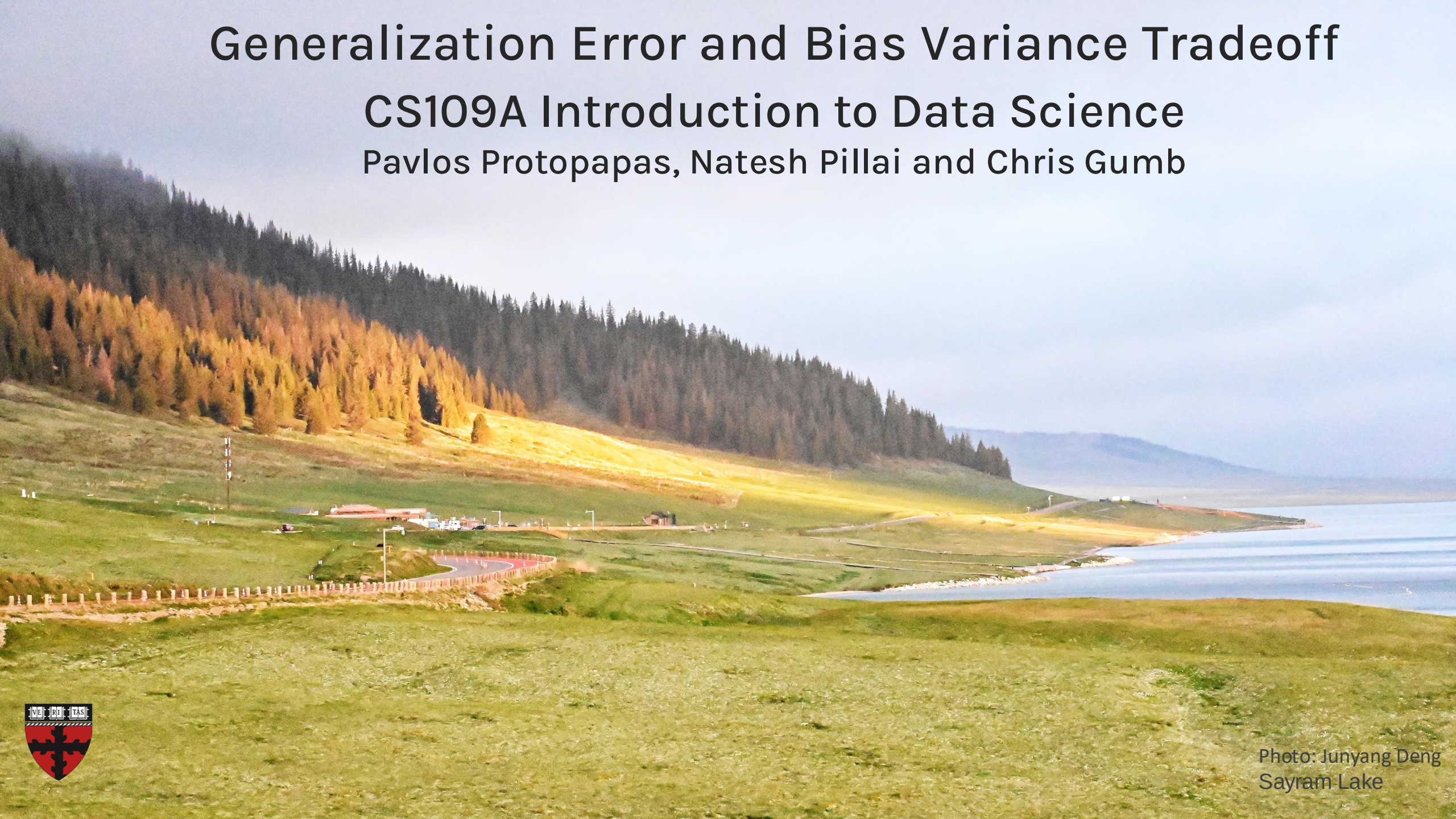


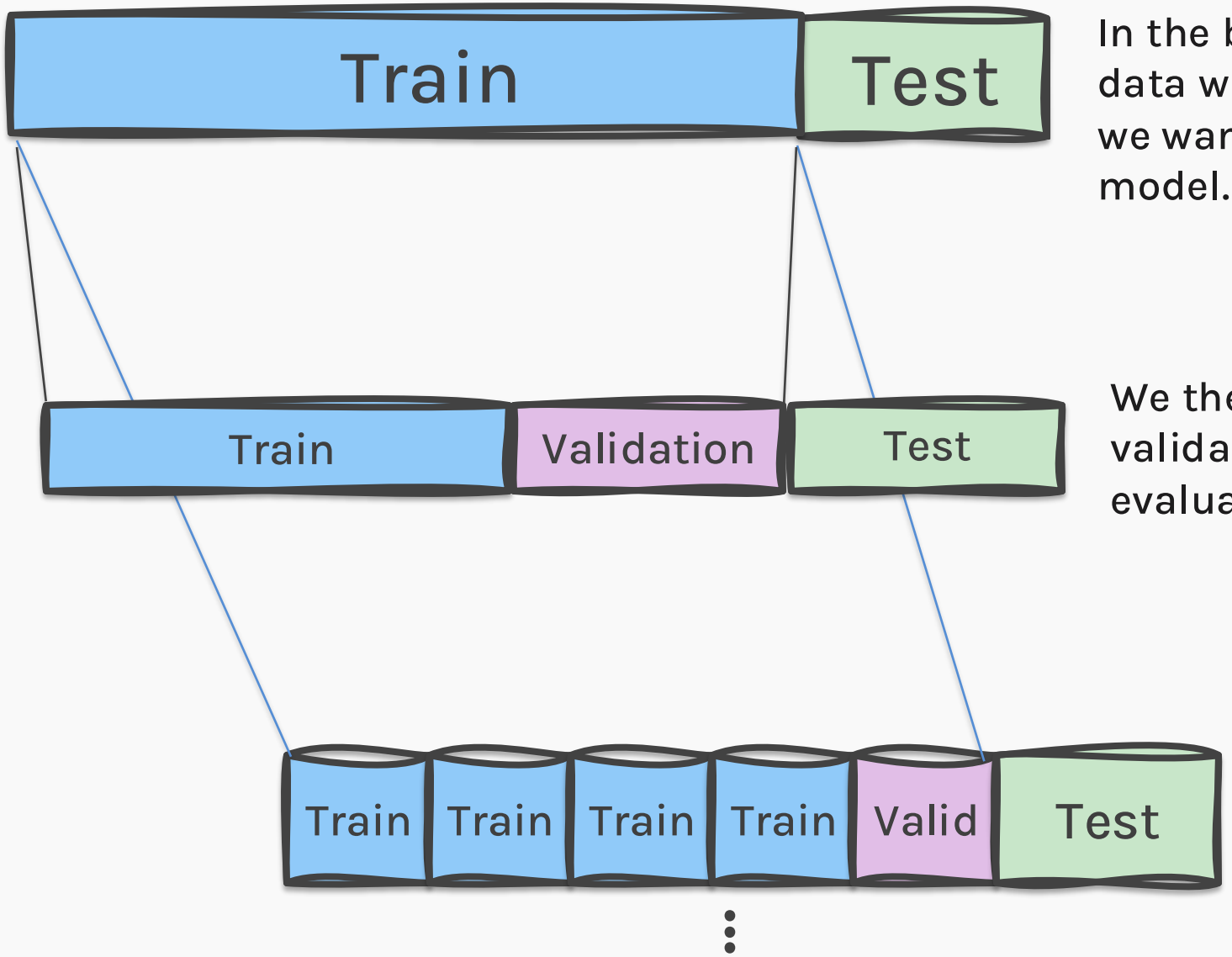
Photo: Junyang Deng
Sayram Lake

Outline

- Recap – Model Selection
- Generalization Error, Bias Variance Tradeoff
- Regularization Techniques: Lasso, Ridge

Outline

- **Recap – Model Selection**
- Generalization Error, Bias Variance Tradeoff
- Regularization Techniques: Lasso, Ridge



In the beginning, we separated a portion of the data which we never touch until the very end when we want to evaluate the performance of the final model. Normally, this is called train + test split. *

We then saw we can split train data into train + validation (to find the best model) + test (to evaluate the performance of the model).

We then finally saw that we can use cross-validation. It splits the train data into k buckets and uses different chunks of data as the validation set.

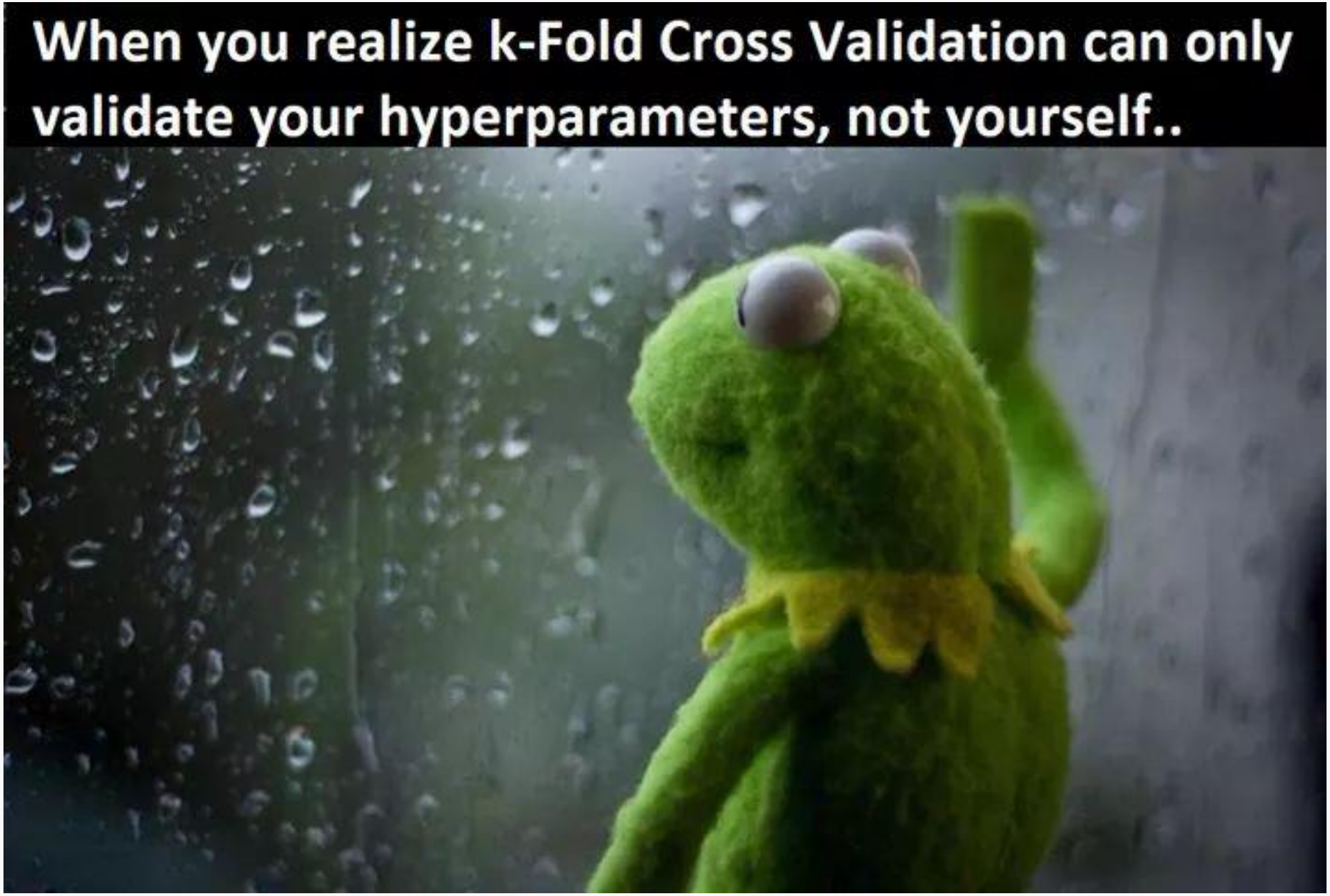
* sometimes they (not us!) also call this train + validation split, while meaning train + test

Recall - Model Selection

1. Model selection as a way to avoid overfitting
2. Validation set to select the best model
3. Cross validation to avoid overfitting to the validation set

Ways of model selection:

- Exhaustive search
- Greedy algorithms
- Fine tuning hyper-parameters
- **Regularization**


When you realize k-Fold Cross Validation can only validate your hyperparameters, not yourself..

Outline

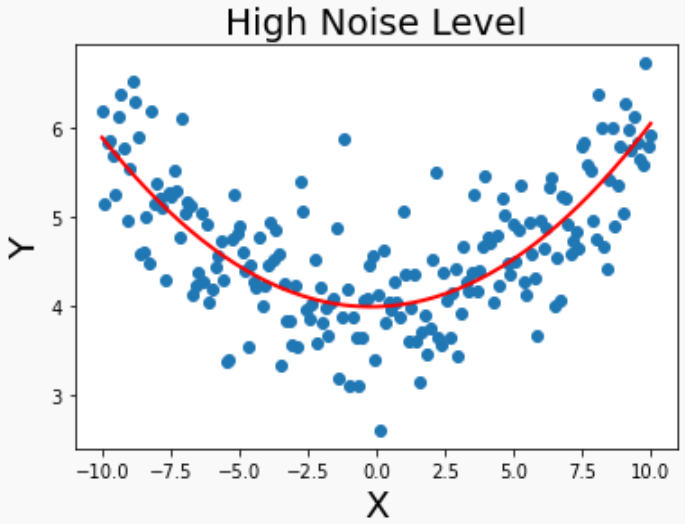
- Recap – Model Selection
- **Generalization Error, Bias Variance Tradeoff**
- Regularization Techniques: Lasso Ridge

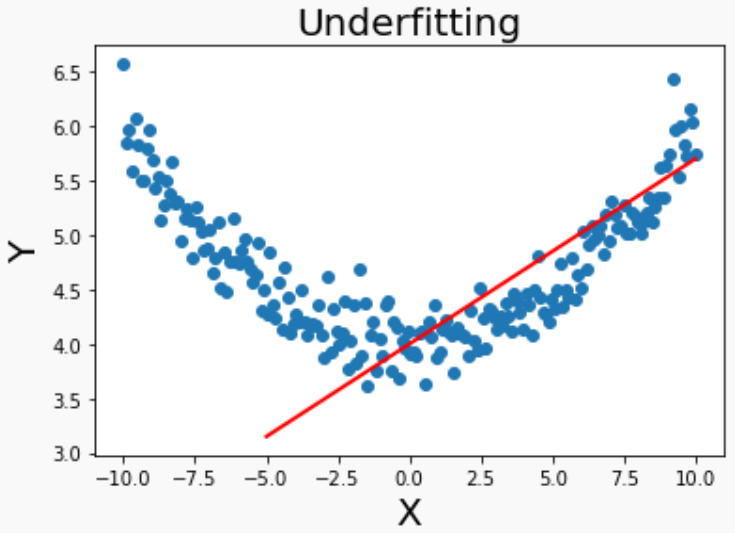
Test Error and Generalization

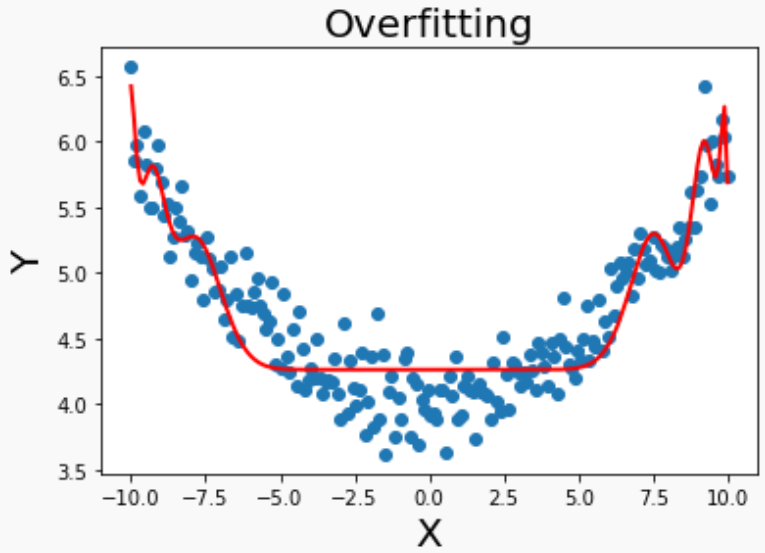
We know to evaluate models on both train and test data because models can do well on train data but do poorly on new data.

When models do well on new data, it is called generalization.

There are at least three ways a model can have a high-test error.







Irreducible and Reducible Errors

We distinguished the contributions of noise to the generalization error:

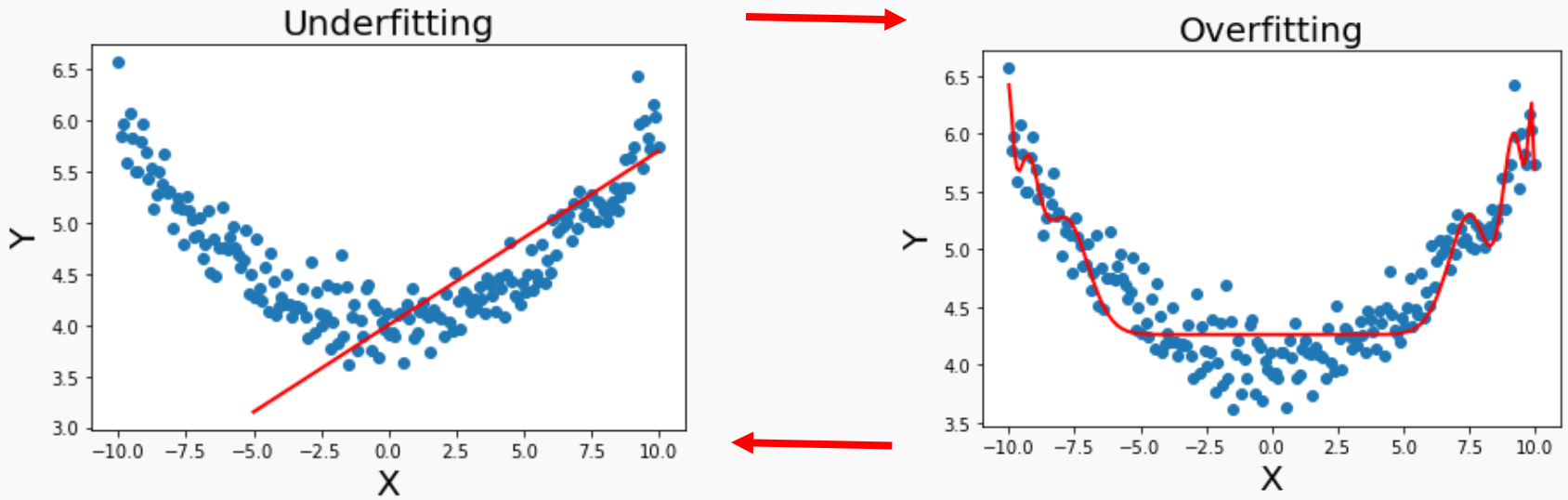
Irreducible error (or aleatoric error): we can't do anything to decrease the error due to noise.

Reducible error (or epistemic error): we can decrease the error due to overfitting and underfitting by improving the model.

The Bias-Variance: Bias

Reducible error comes from either underfitting or overfitting. There is a tradeoff between the two sources of errors:

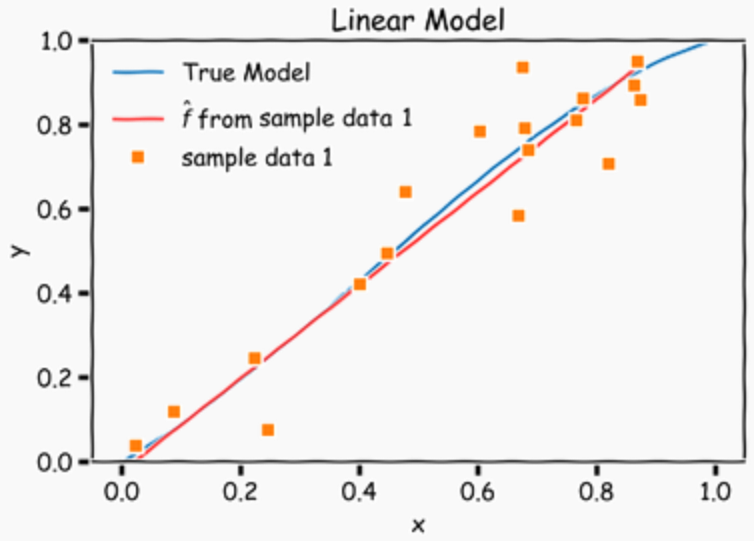
Increase complexity

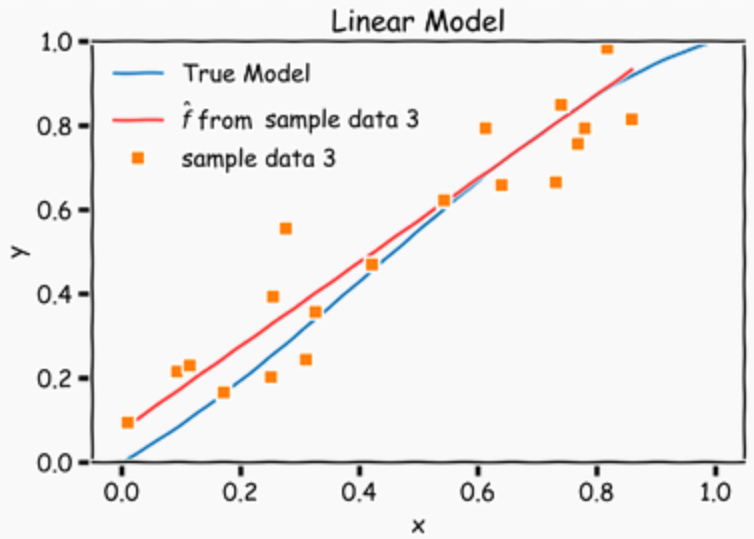


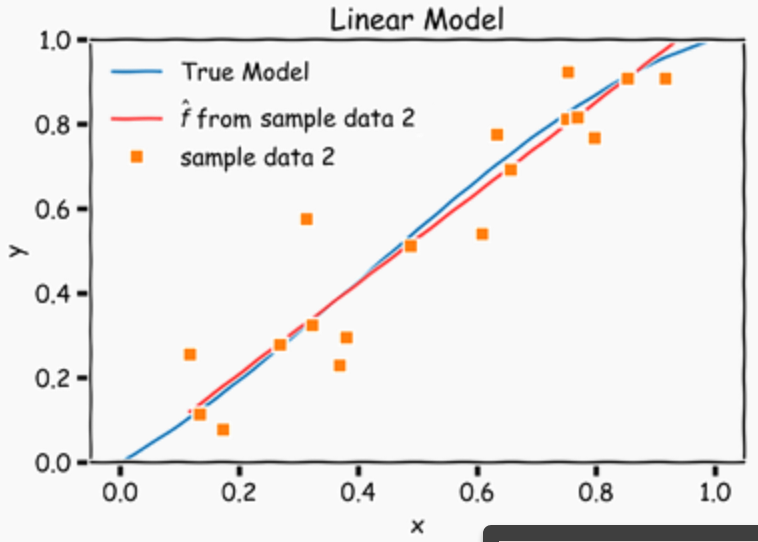
Decrease complexity

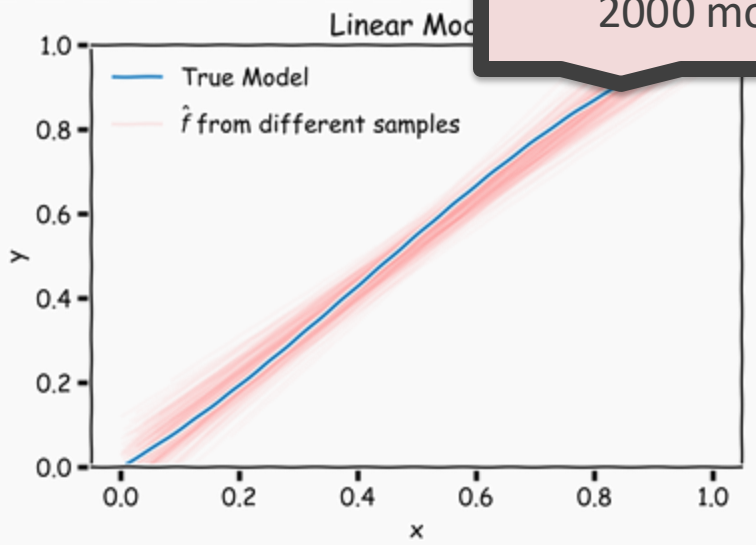

NYPD
WITH GREAT
MODEL COMPLEXITY
COMES GREAT
RISK OF OVERFITTING
imgflip.com

Bias vs Variance: Variance of a SIMPLE model



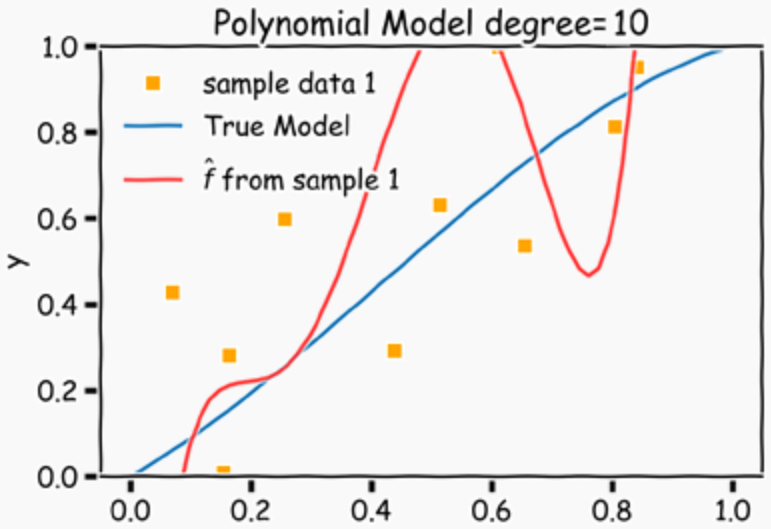


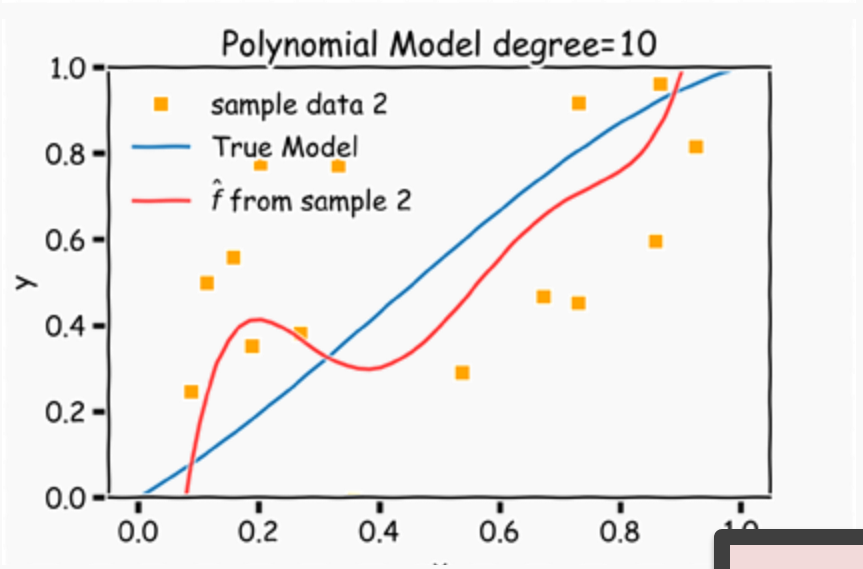


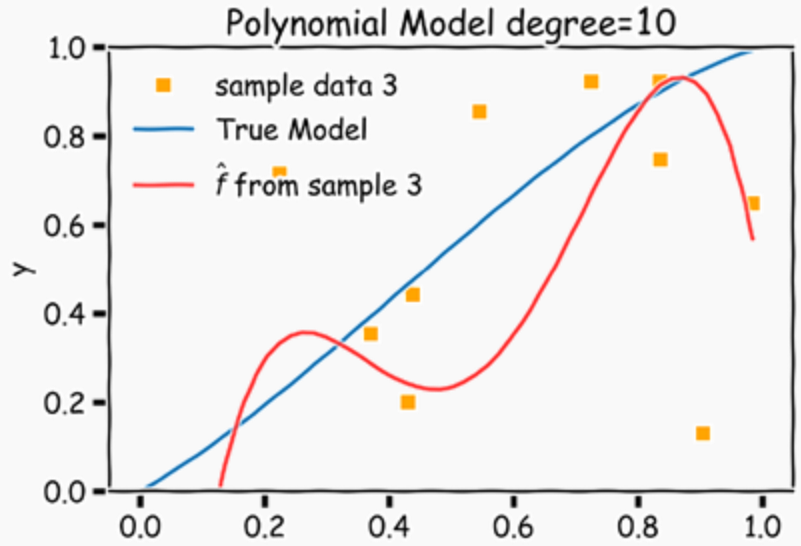


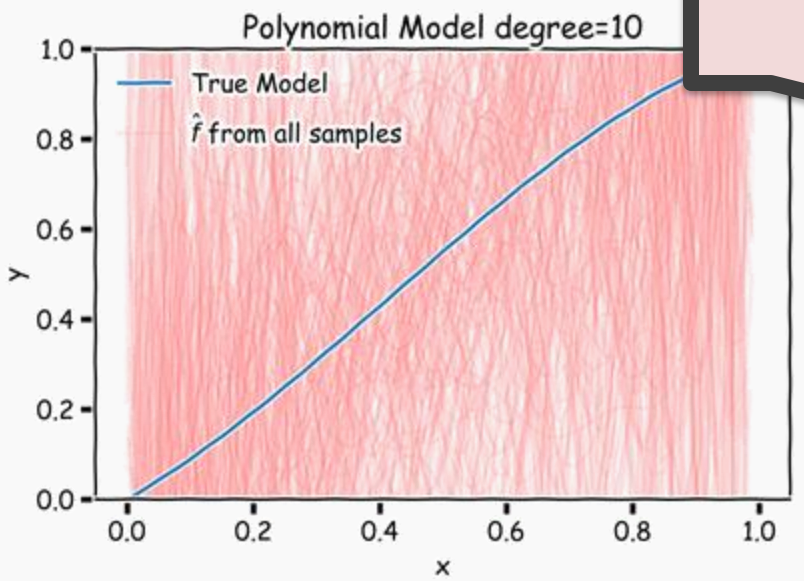
2000 models

Bias vs Variance: Variance of a COMPLEX model





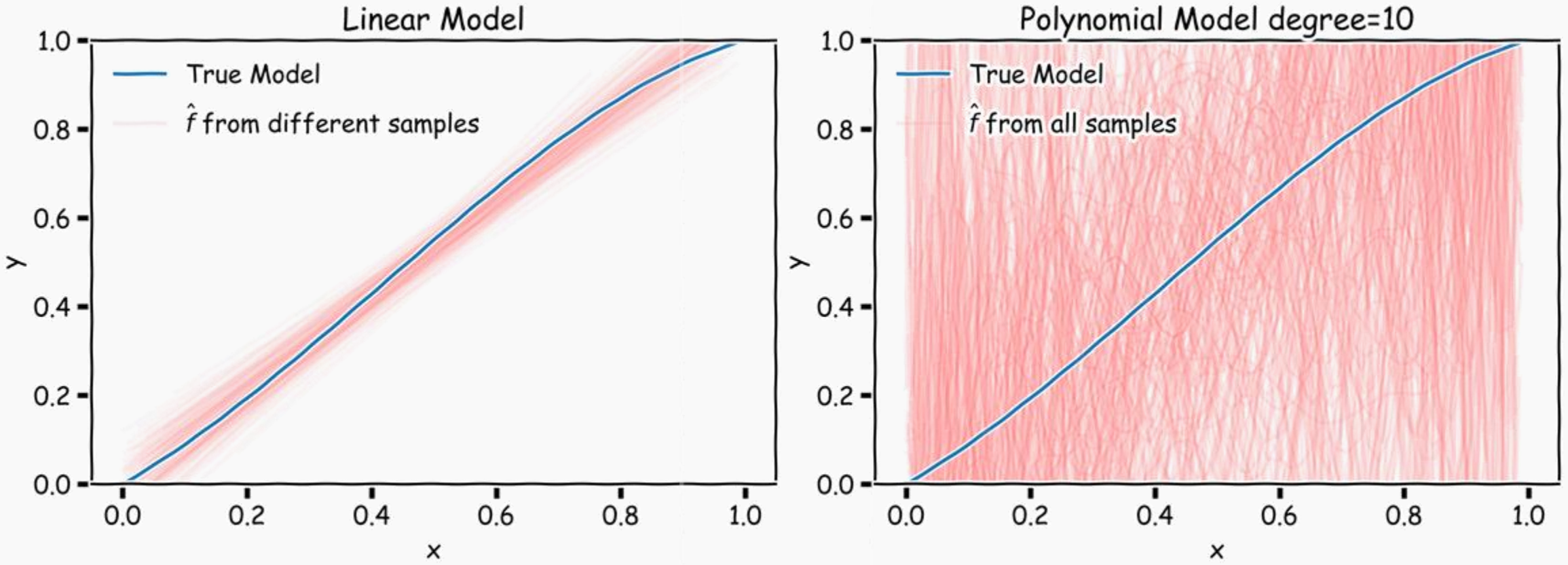




2000 models

Bias vs Variance

Left: 2,000 best-fit linear models, each fitted to a different 20-point training set.
Right: 2,000 best-fit models using degree-10 polynomials.

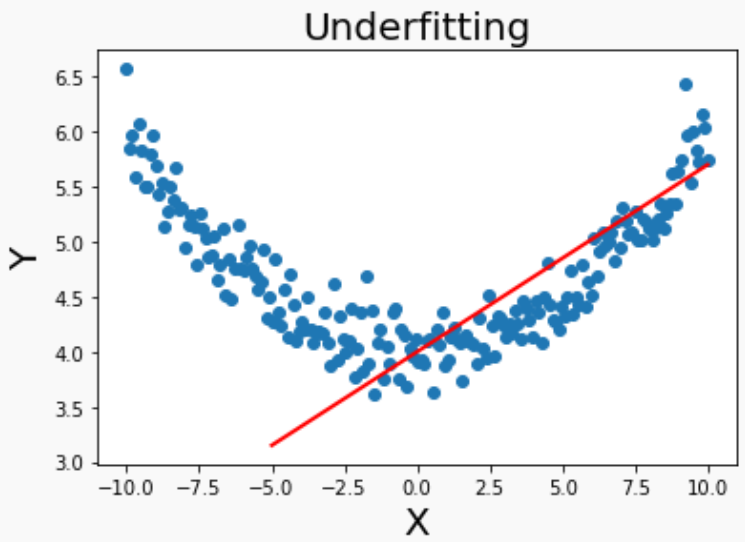


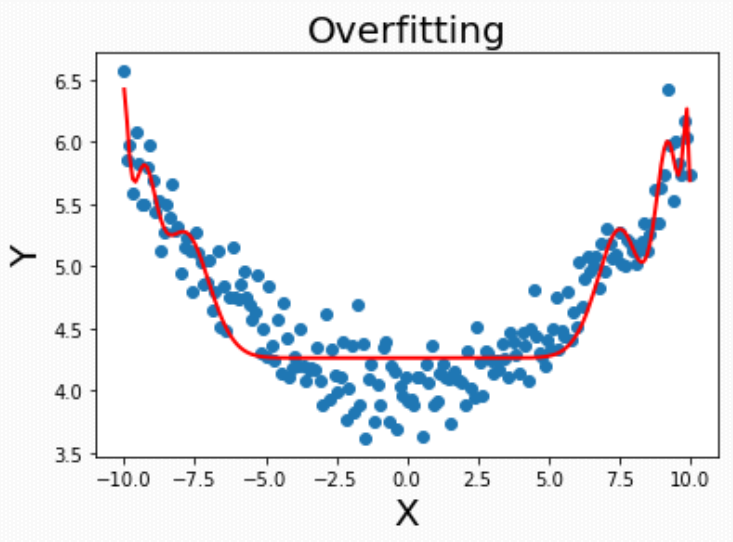
The Bias-Variance: Bias

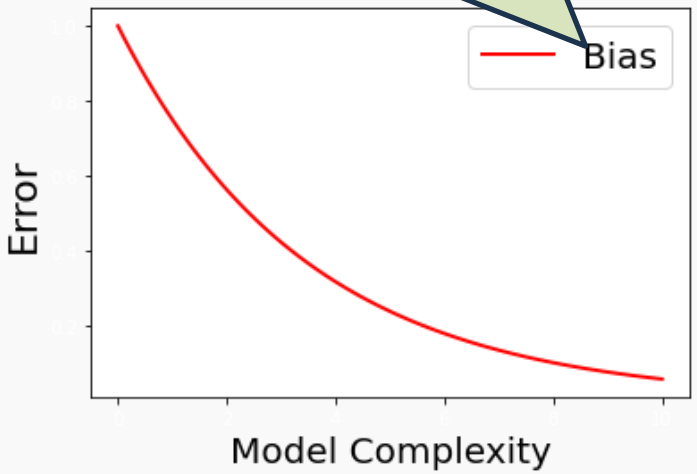
Reducible error comes from either underfitting or overfitting. There is a tradeoff between the two sources of errors:

"bias" refers to how far off a model's predictions are from the actual truth.

Increase complexity



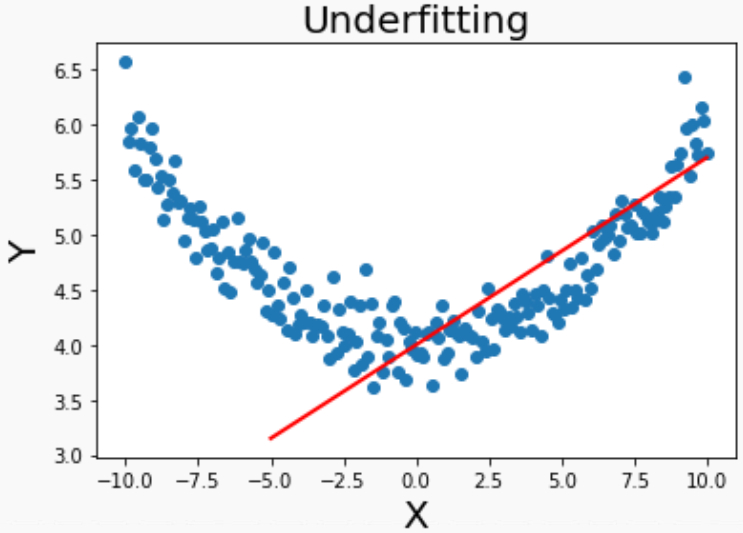


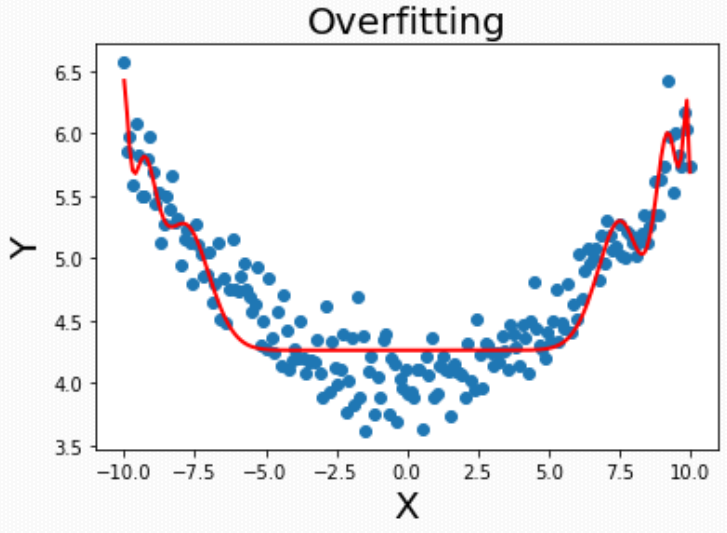


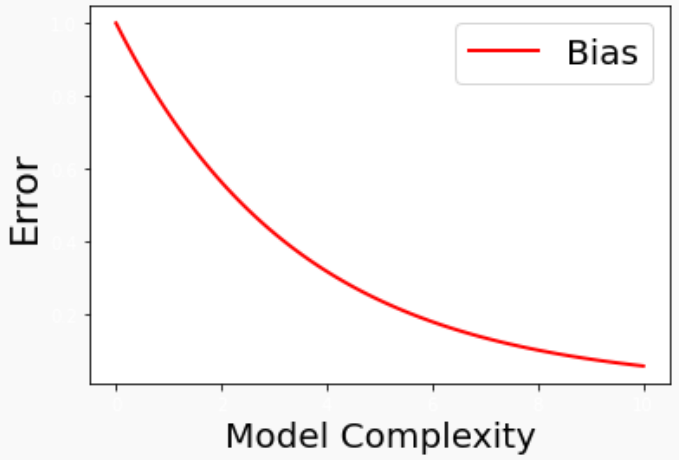
Decrease complexity

The Bias-Variance Trade Off

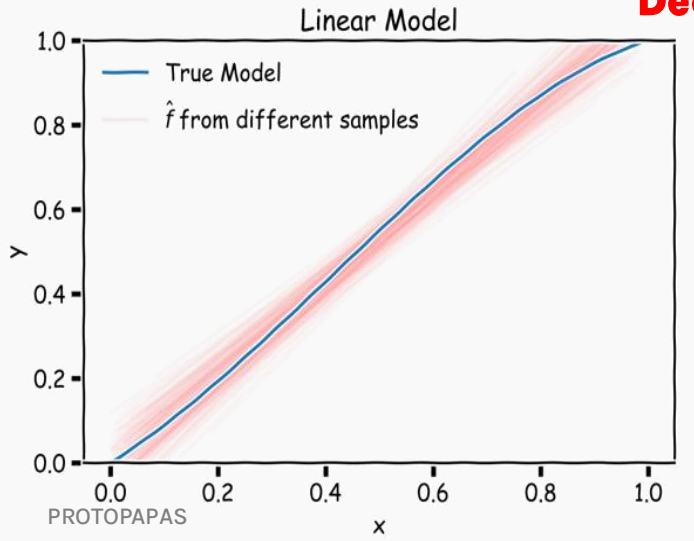
Increase complexity

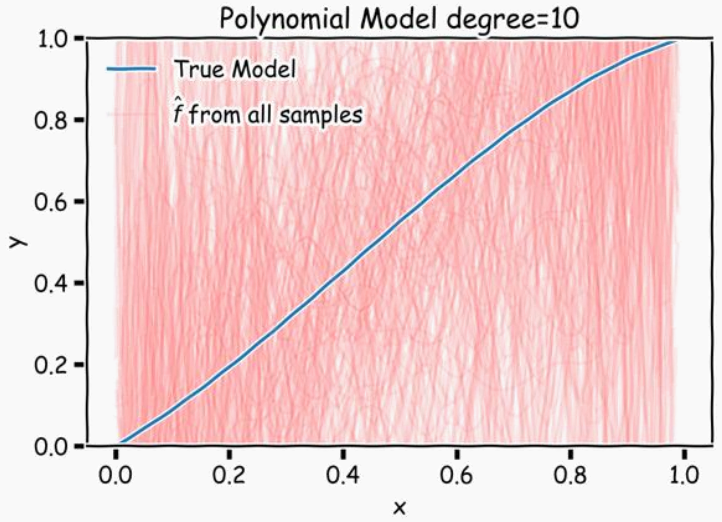


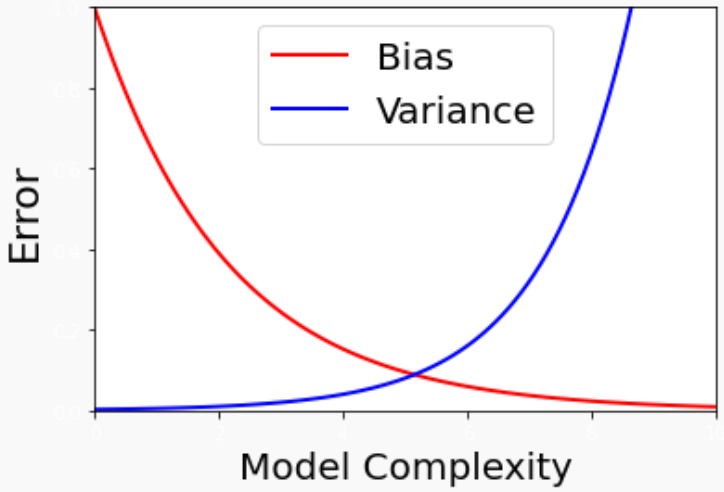




Decrease complexity

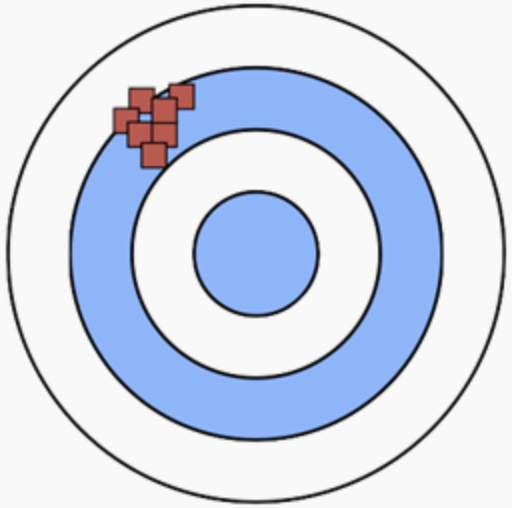


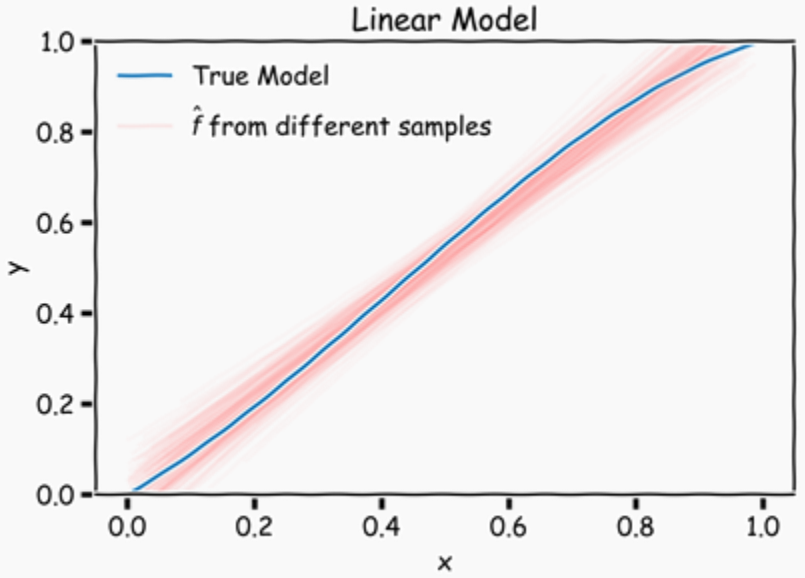




Low Variance
(Precise)

High Bias
(Not Accurate)

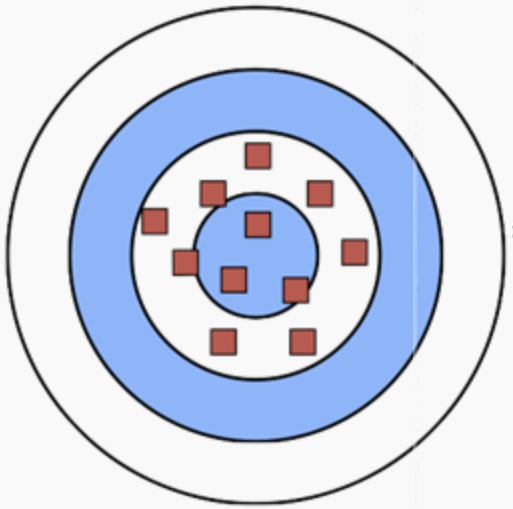


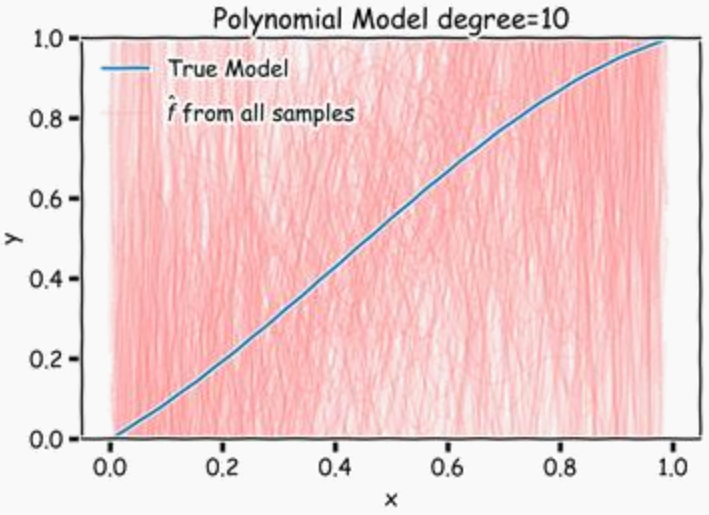


Low Variance
(Precise)

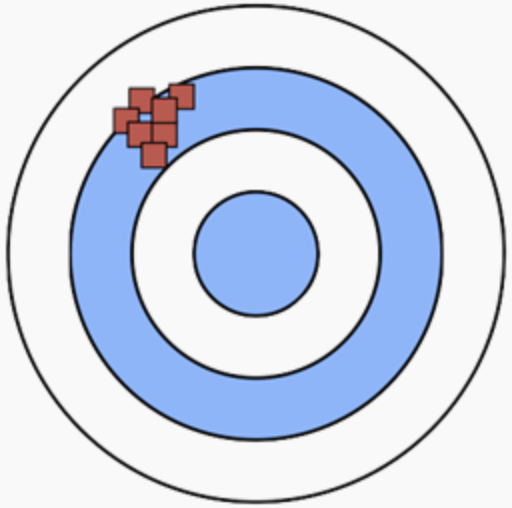
High Variance
(Not Precise)

Low Bias
(Accurate)

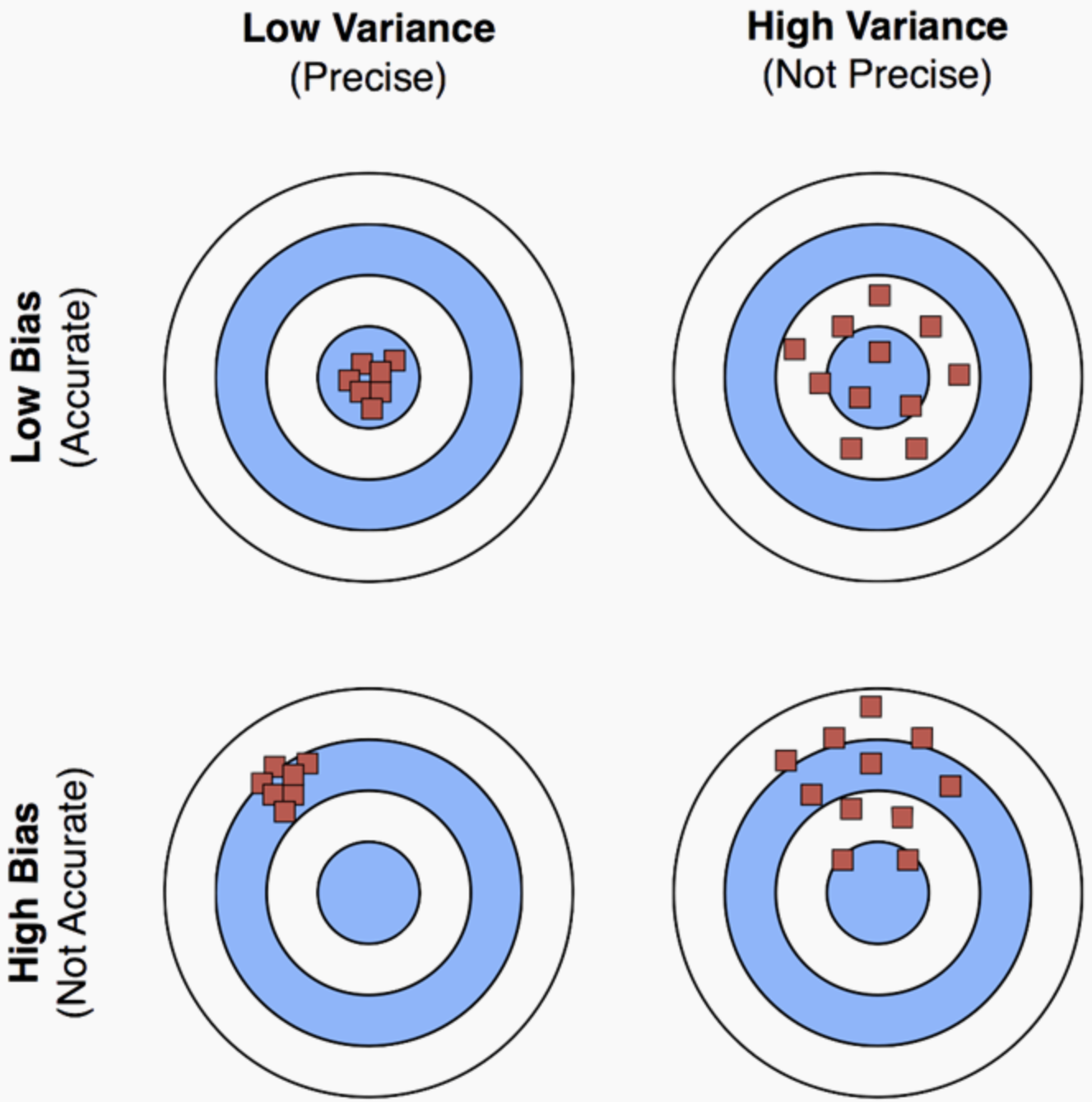




High Bias
(Not Accurate)



WE WANT THIS



WE WANT TO AVOID THIS

Overfitting

Overfitting occurs when a model corresponds too closely to the training set, and as a result, the model fails to fit additional data.

So far, we have seen that overfitting can happen when:

- too many parameters
- the degree of the polynomial is too large
- too many interaction terms

Soon, we will see other evidence of overfitting, which will point to a way of avoiding overfitting: Ridge and Lasso regressions.