GENERAL ASSEMBLY

EVALUATING MODEL FIT

Gus Ostow

Machine Learning Consultant, Papert Labs

LEARNING OBJECTIVES

- Evaluate regression fit with error metrics
- Understand the tradeoff between under-fitting and over-fitting, and how that is impacted by decisions about training data and hyper-parameter choice.
- Justify when regularization is a smart choice for a linear model

PRE-WORK REVIEW

- Understand goodness of fit (r-squared)
- Measure statistical significance of features
- Recall what a residual is
- Implement a sklearn estimator to predict a target variable

OPENING

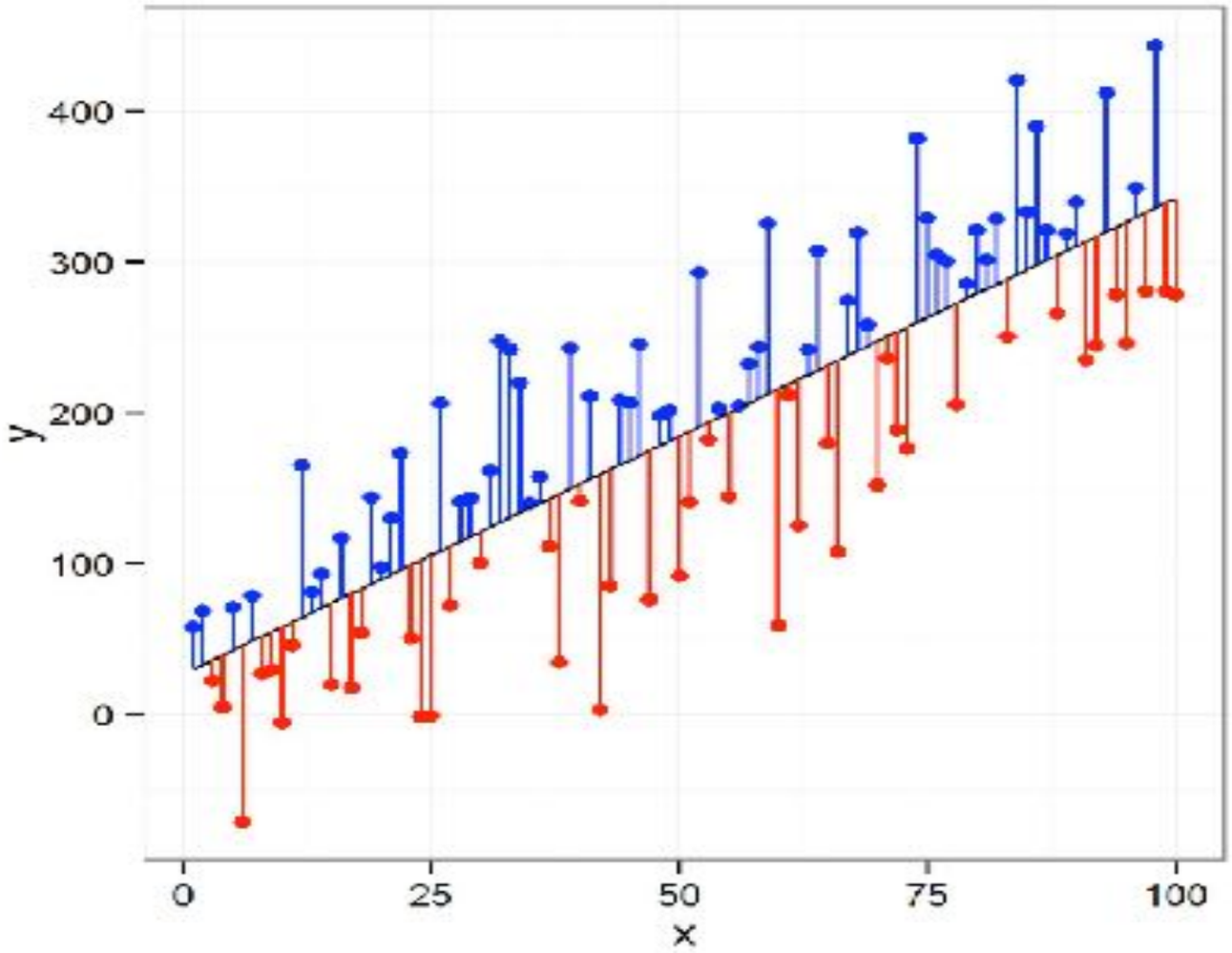
R-SQUARES AND RESIDUALS

WHAT IS R-SQUARED? WHAT IS A RESIDUAL?

- R-squared, the central metric introduced for linear regression
- Which model performed better, one with an r-squared of 0.79 or 0.81?
- R-squared measures explain variance.
- But does it tell the magnitude or scale of error? It isn't super informative from a business perspective
- We'll review the unscaled loss functions from yesterday.

LINEAR MODELS AND ERROR

RECALL: WHAT'S RESIDUAL ERROR?



MEAN SQUARED ERROR (MSE)

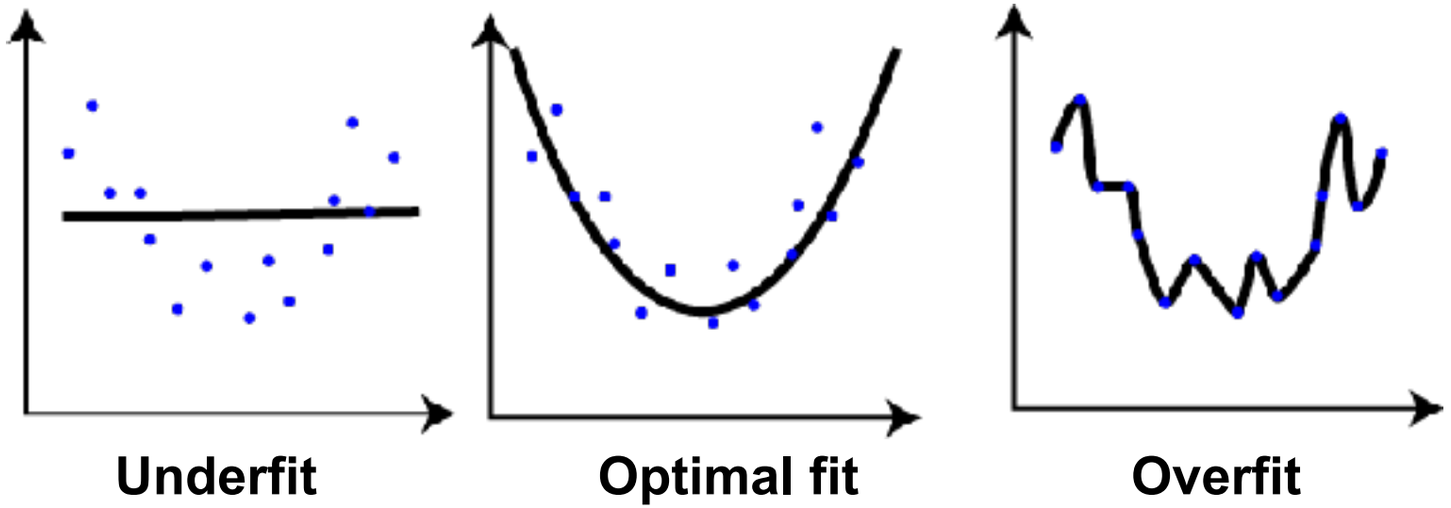
- To calculate MSE:
  - Calculate the difference between each target y and the model's predicted value y-hat (i.e. the residual)
  - Square each residual.
  - Take the mean of the squared residual errors.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

MODEL FIT

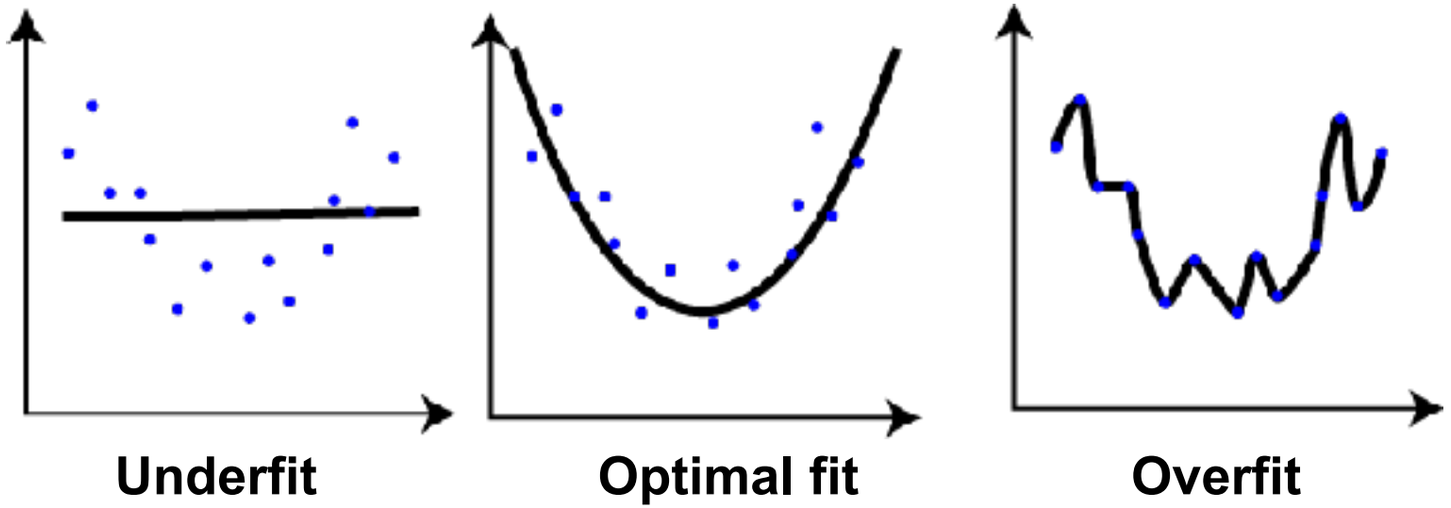
OVERFITTING vs UNDERFITTING

WHAT IS OVERFITTING?



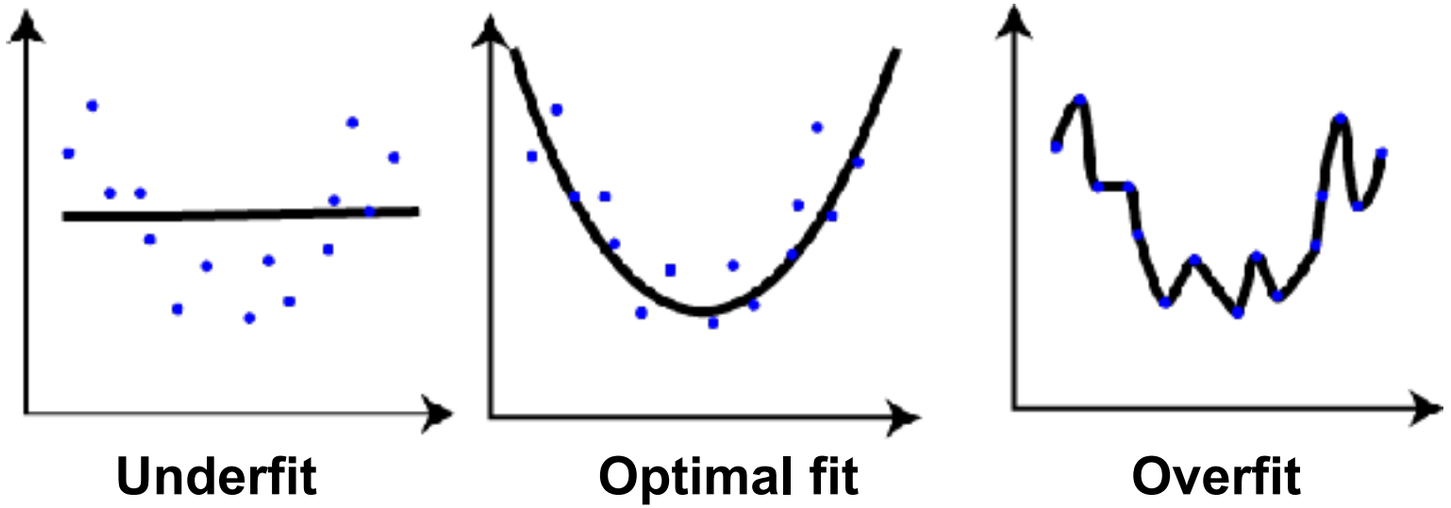
- Is not picking up any pattern in the data
- The second model explains the general curve of the data.
- The third model drastically overfits the model, bending to every point.

WHAT IS OVERFITTING?



- Is not picking up any pattern in the data
- The second model explains the general curve of the data.
- The third model drastically overfits the model, bending to every point.

WHAT IS OVERFITTING?



Underfit model

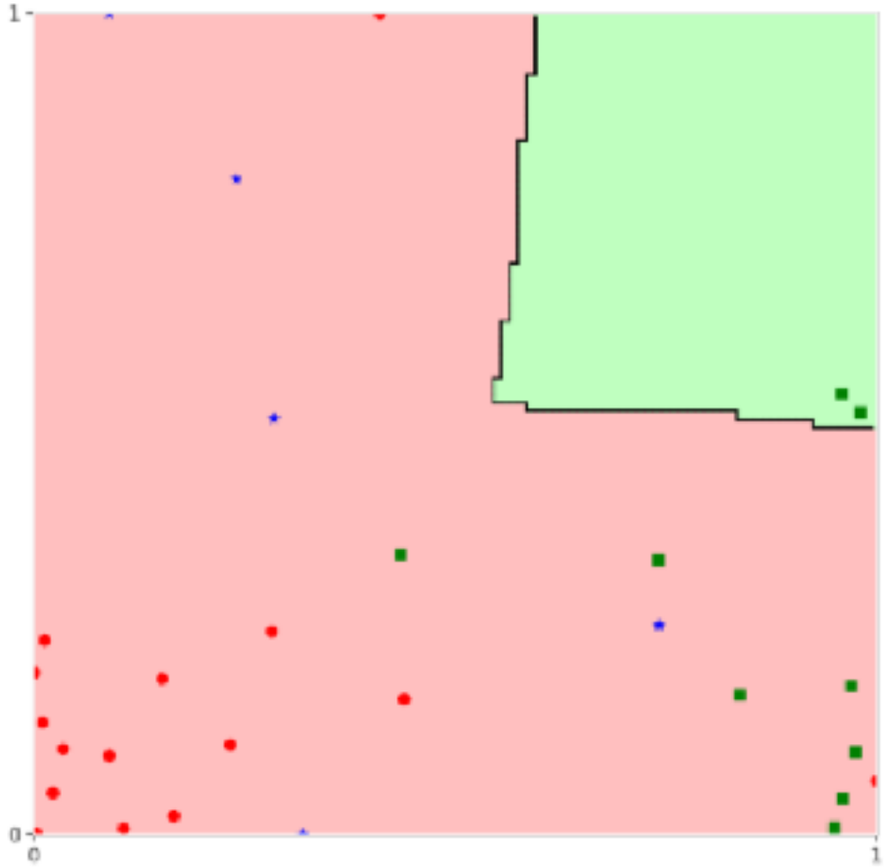
- Low score on training set
- Low-ish score on test set

Overfit model

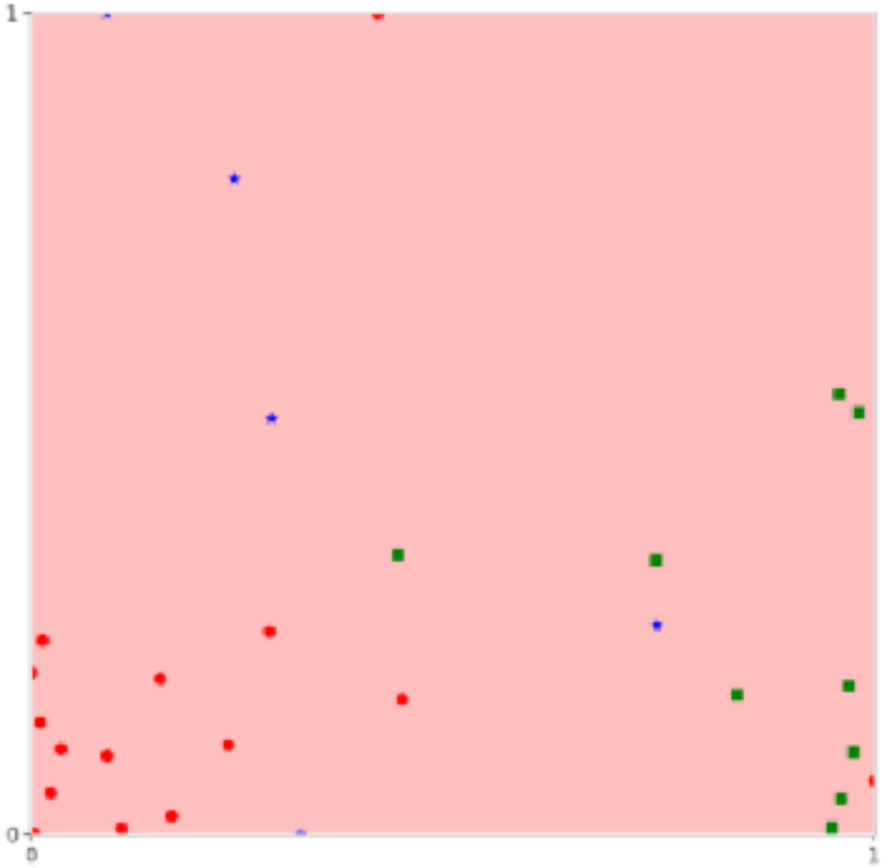
- Great score on training set
- Horrible score on test set

KNN Fit Case-Study

$k = 25$

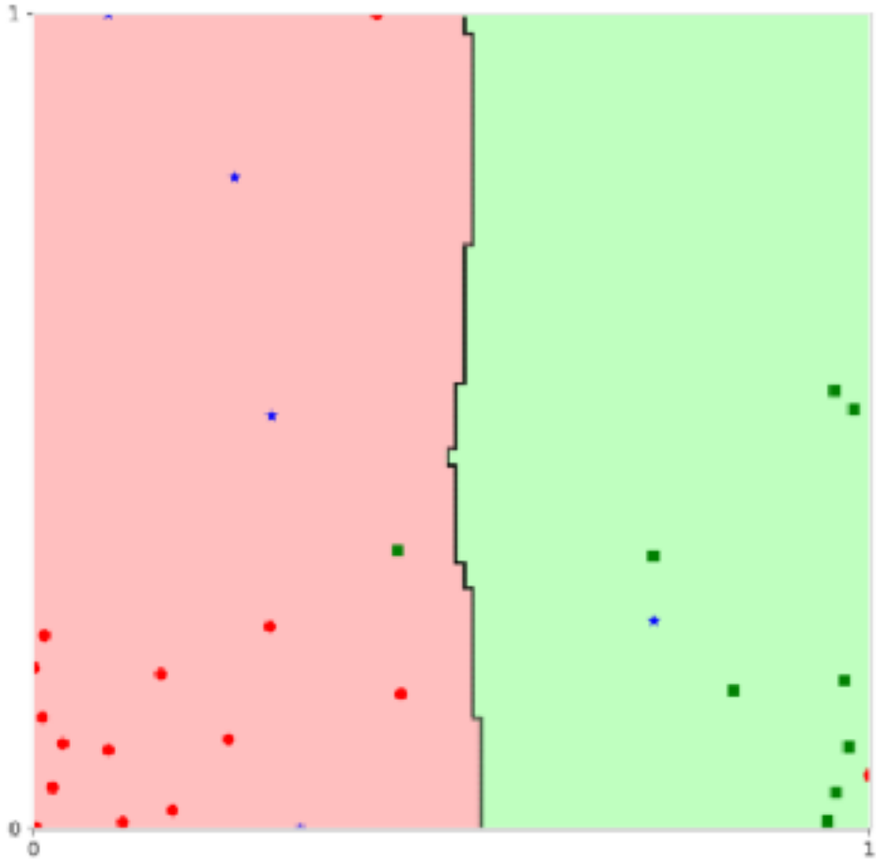


$k = 26$

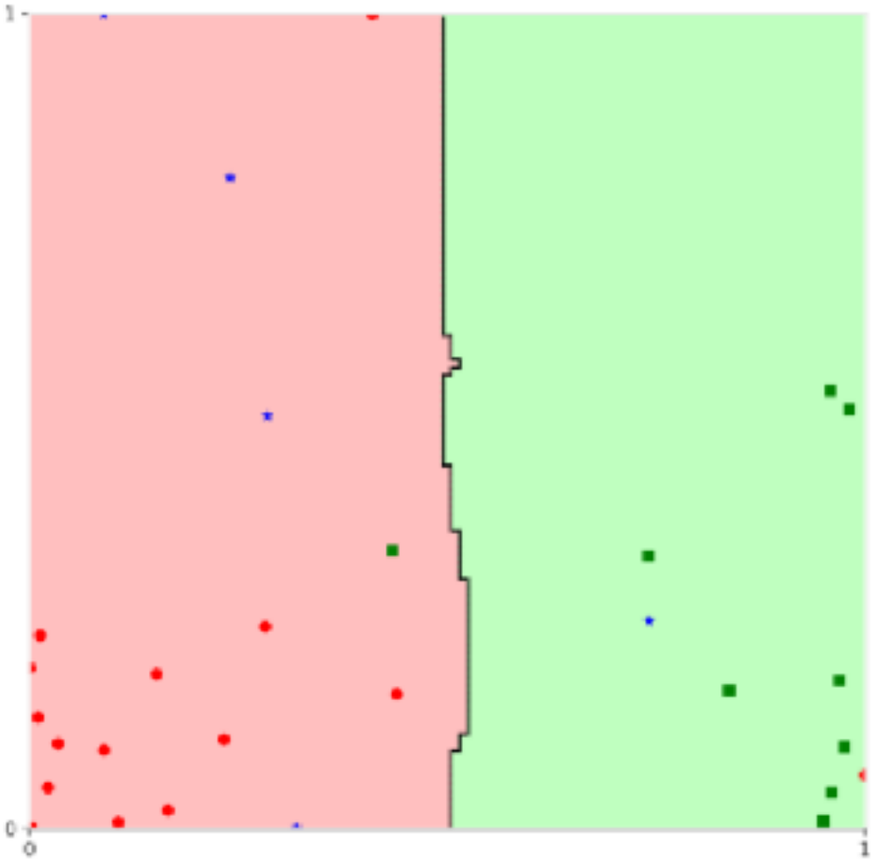


KNN Fit Case-Study

$k = 15$

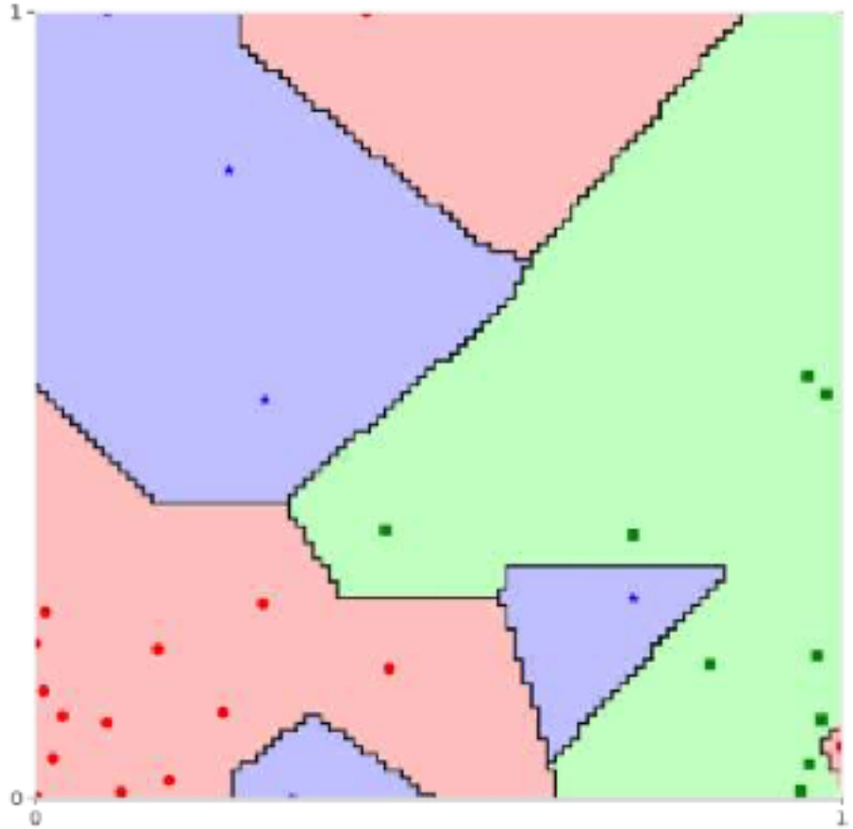


$k = 16$

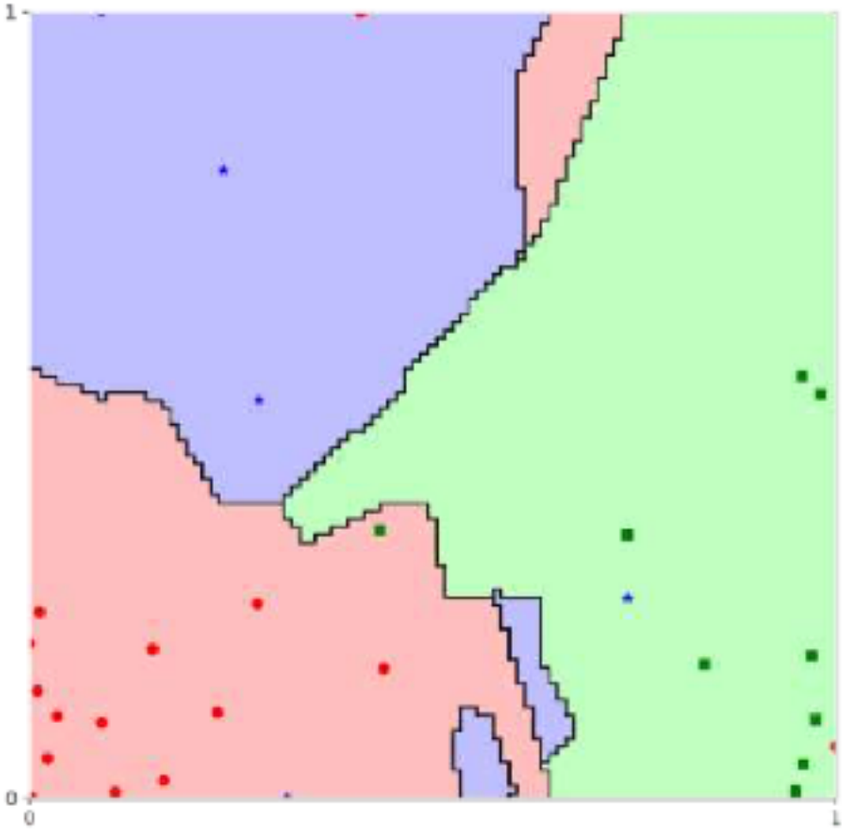


KNN Fit Case-Study

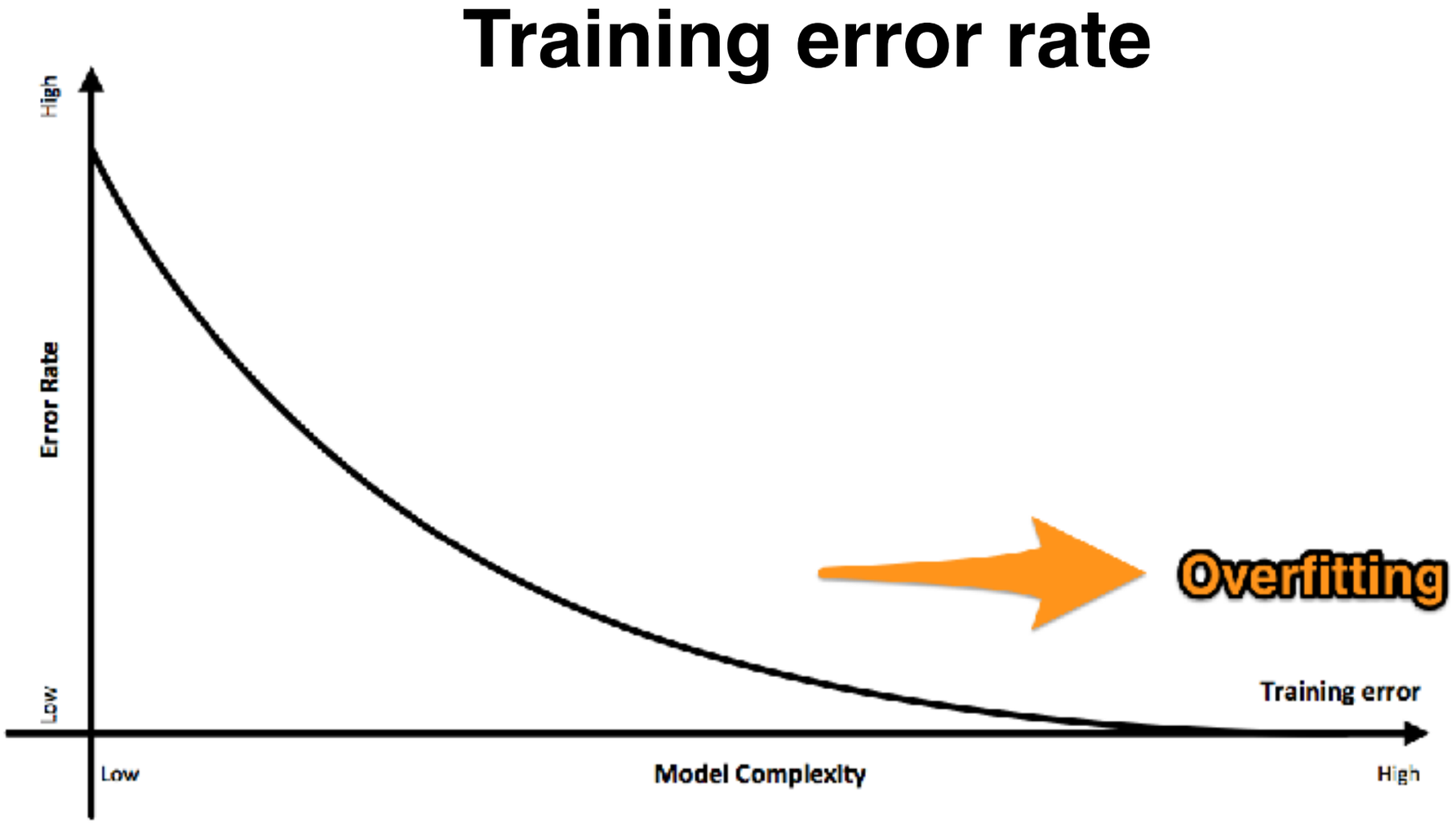
$k = 2$



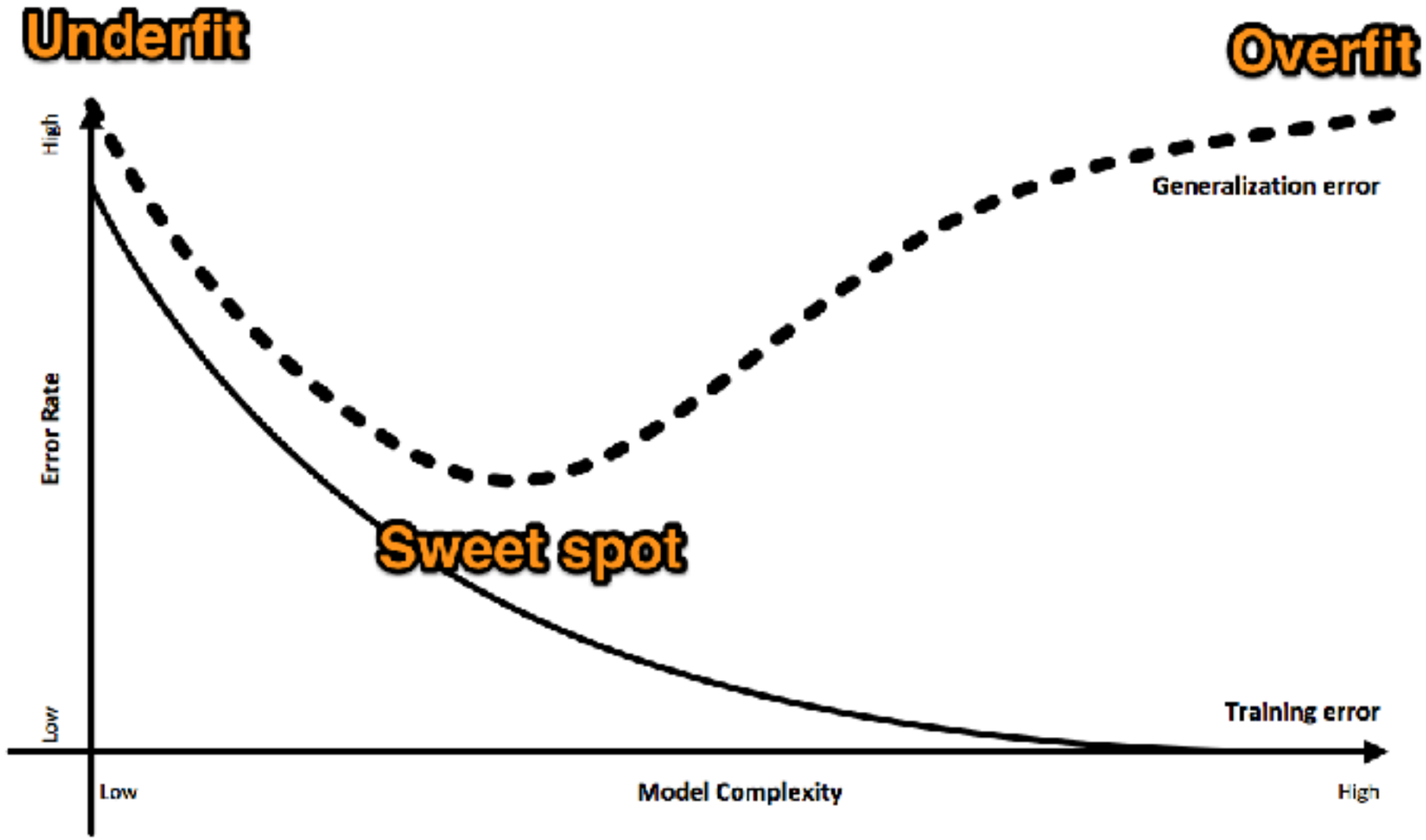
$k = 3$



“Complex models overfit”

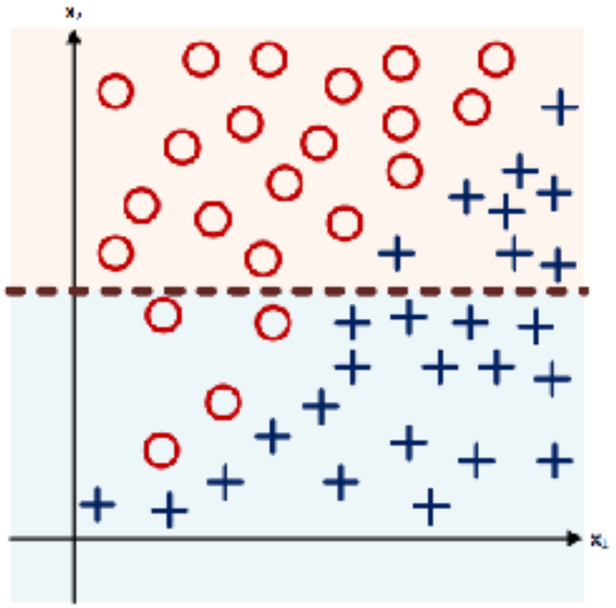


“Complex models overfit”

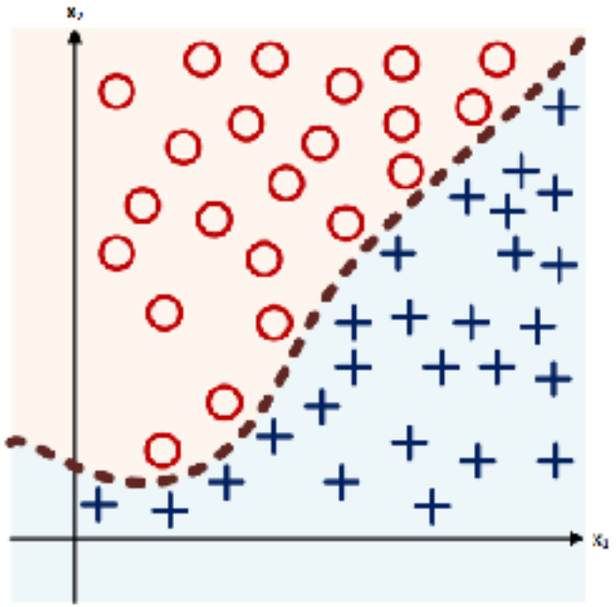


TYPES OF FIT

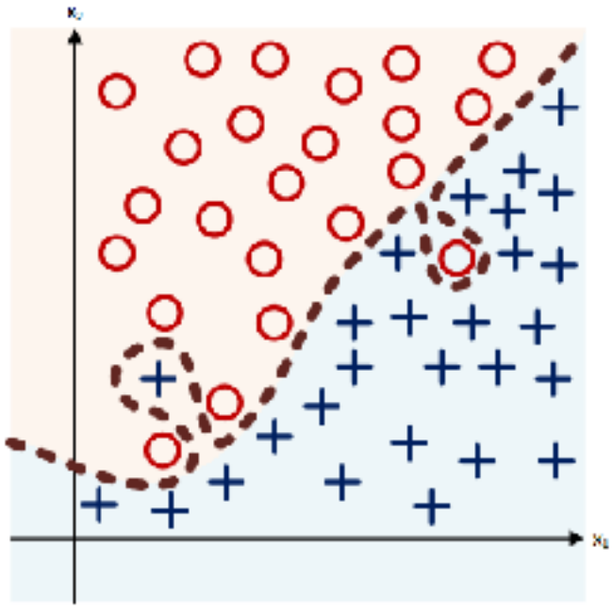
Underfit



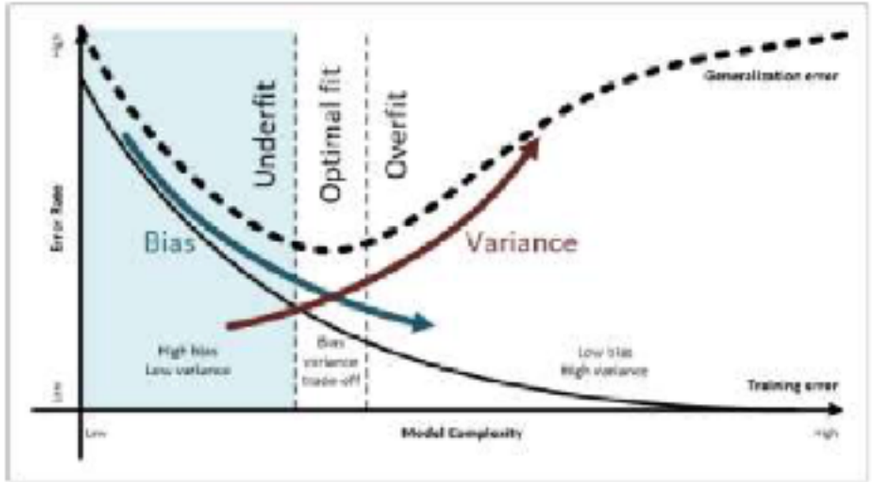
Optimal Fit

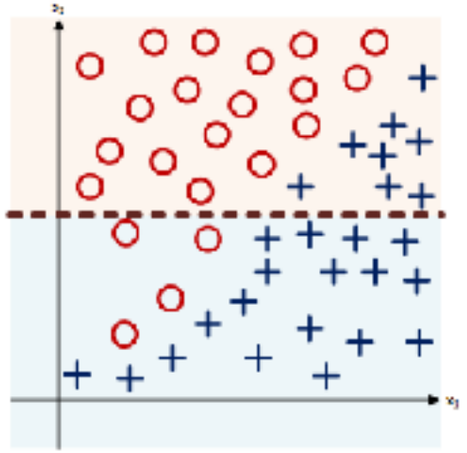


Overfit



UNDERFIT

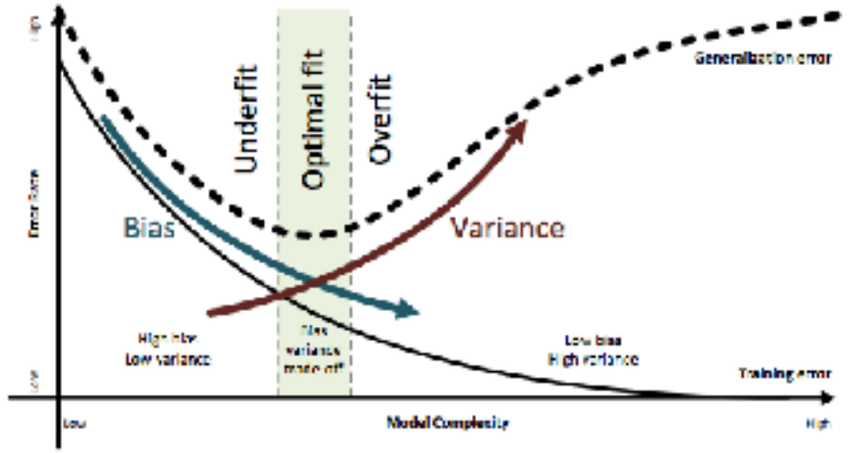


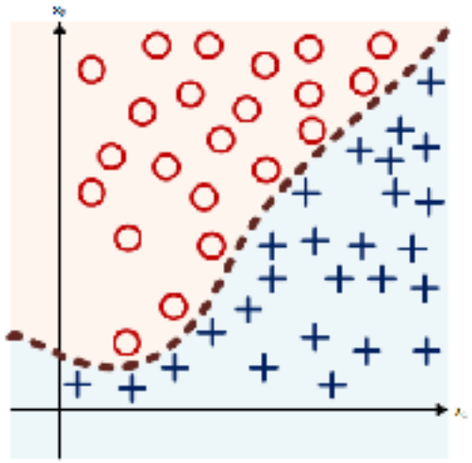


- Underfit
  - Model too simple
  - It cannot represent the desired behavior very well; both its training and generalization error are poor
  - High bias; low variance

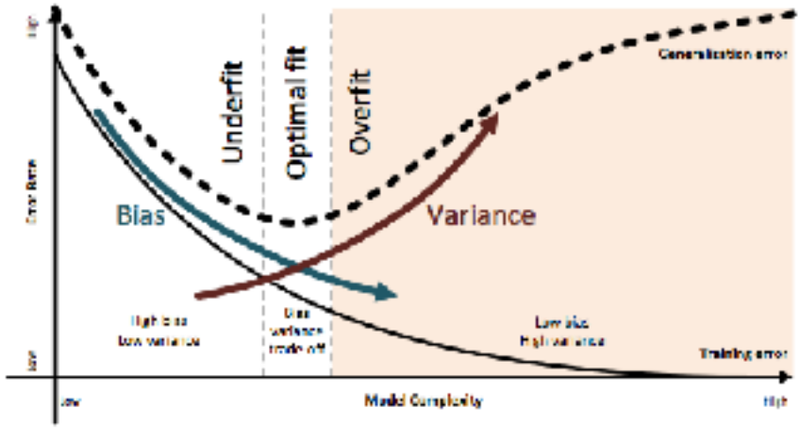
OPTIMAL FIT

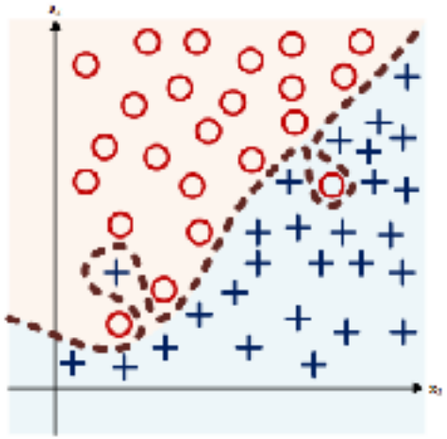
- Optimal Fit
  - Model has the right level of complexity
  - It performs well on the training set (low training error) and generalize well to unknown data points (low generalization error)





OVERFIT





- Overfit
  - Model too complex
  - It performs very well on the training set (low training error) but does not generalize well to unseen data points (high generalization error)
  - Low bias; high variance

ACTIVITY: KNOWLEDGE CHECK

ANSWER THE FOLLOWING QUESTIONS (5 minutes)

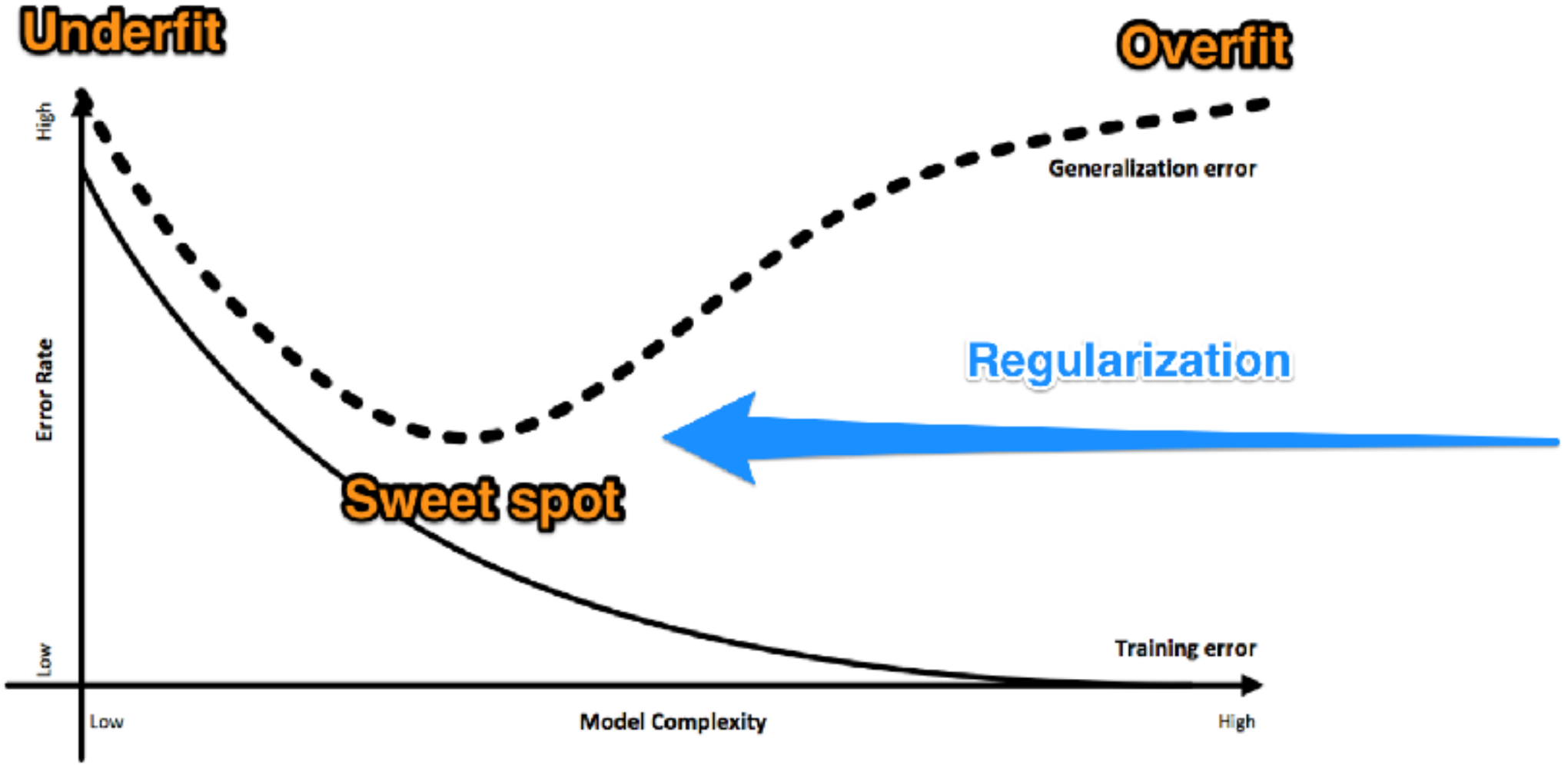
1. Which of the following scenarios would be better for a weatherman?
   a. Knowing that I can very accurately "predict" the temperature outside from previous days perfectly, but be 20-30 degrees off for future days
   b. Knowing that I can accurately predict the general trend of the temperate outside from previous days, and therefore am at most only 10 degrees off on future days

DELIVERABLE

Answers to the above questions

REGULARIZATION

“Complex models overfit”



WHAT IS REGULARIZATION? AND WHY DO WE USE IT?

- Regularization is an additive approach to protect models against overfitting (being potentially biased and overconfident, not generalizing well).
- Regularization becomes an additional weight to coefficients, shrinking them closer to zero.
- L1 (Lasso Regression) adds the extra weight to coefficients.
- L2 (Ridge Regression) adds the square of the extra weight to coefficients.
- Use Lasso when we have more features than observations ($k > n$) and Ridge otherwise.

WHERE REGULARIZATION MAKES SENSE

- It doesn’t seem to help. Why is that?

- We need to optimize the regularization weight parameter (called alpha) through cross validation.

LESSON REVIEW

- What's the (typical) range of r-squared?
- What's the range of mean squared error?
- How would changing the scale or interpretation of y (your target variable) effect mean squared error?
- What are the two main contributors to your model’s poor performance?
- How do you diagnose if you are overfitting,underfitting, or killing the game?