

CHAPTER 2: SUPERVISED LEARNING

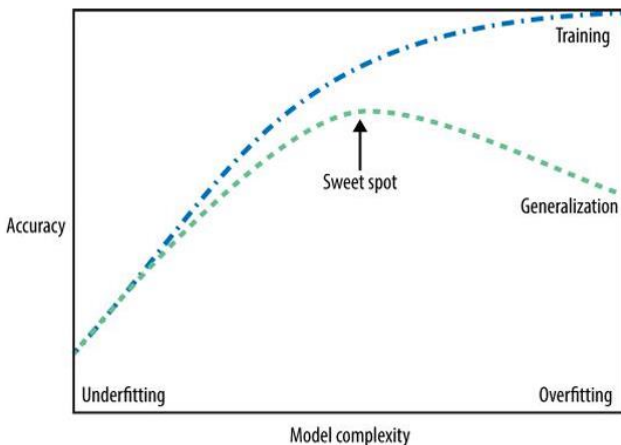
- Supervised learning is used whenever we want to predict a certain outcome from a given input, and we have examples of input/output pairs.

CLASSIFICATION AND REGRESSION

- Two major types of supervised learning: Classification and Regression.
- **Classification:**
Goal: Predict a class label, which is a choice from a predefined list of possibilities (labels).
Binary Classification: two classes. Yes/no questions.
Example: Detecting spam emails.
Multiclass Classification: more than two classes.
Example: Iris species classification.
- **Regression:**
Goal: Predict a real/continuous number.
Example: Predict a person's annual income from their education, age, where they live,...
Predict the yield of a corn farm from previous yields, weather, employees,...

GENERALIZATION, OVERFITTING, UNDERFITTING

- If a model is able to make accurate predictions on unseen data, it is able to *generalize* from the training set to the test set.
- **Overfitting** is when you fit a model too closely to the particularities of the training set and obtain a model that works well on the training set but is not able to generalize to new data.
- **Underfitting** is choosing too simple of a model.



RELATION OF MODEL COMPLEXITY TO DATASET SIZE

- Having more data and building appropriately more complex models can often work wonders for supervised learning tasks.

SUPERVISED LEARNING ALGORITHMS

K-Nearest Neighbors

- To make a prediction for a new data point, the algorithm finds the closest data points in the training datasets – “nearest neighbors”.

K-Neighbors classification

-