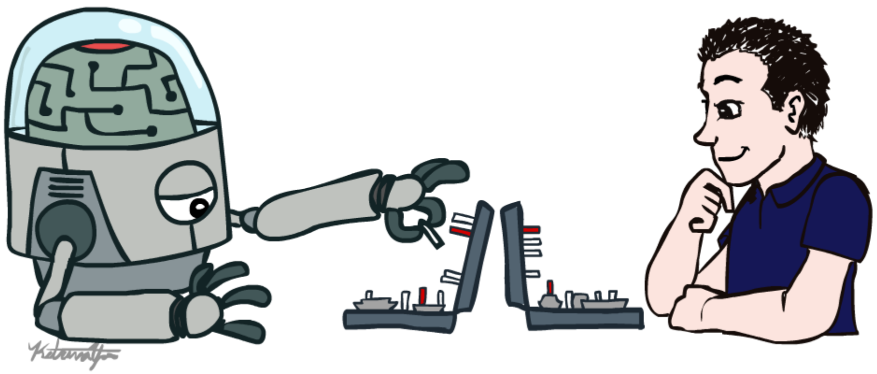
CSE 3521: Introduction to Artificial Intelligence



[Many slides are adapted from the UC Berkeley. CS188 Intro to AI at UC Berkeley and previous CSE 3521 course at OSU.]

THE OHIO STATE UNIVERSITY

Machine learning

A set of methods that can automatically <u>detect patterns</u> in data, and then use the <u>uncovered patterns to predict</u> future data, or to perform other kinds of <u>decision making under uncertainty</u>

Kevin Murphy. Machine learning: a probabilistic perspective. The MIT Press

Learning from Data

Algorithm | **Evaluation** | **Data**

Linear regression | SSE | $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$

Least square (closed-form)

Supervised learning

- Data type: $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$

- Goal: Build a model so that given a future data instance $\boldsymbol{x}$, it can tell the label y
    - Example: Nearest neighbors

- The "label" in $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$ provides supervision of how to give each data instance a label
- The label can be **"numerical" (regression)** or **"categorical" (classification)**

Classification vs. regression

- Classification
  - Supervised learning
  - Form: Data point → Desired category (integer number index)
  - Ex: 1: Cat, 2: Dog, 3: Horse, ……, 1000: Car

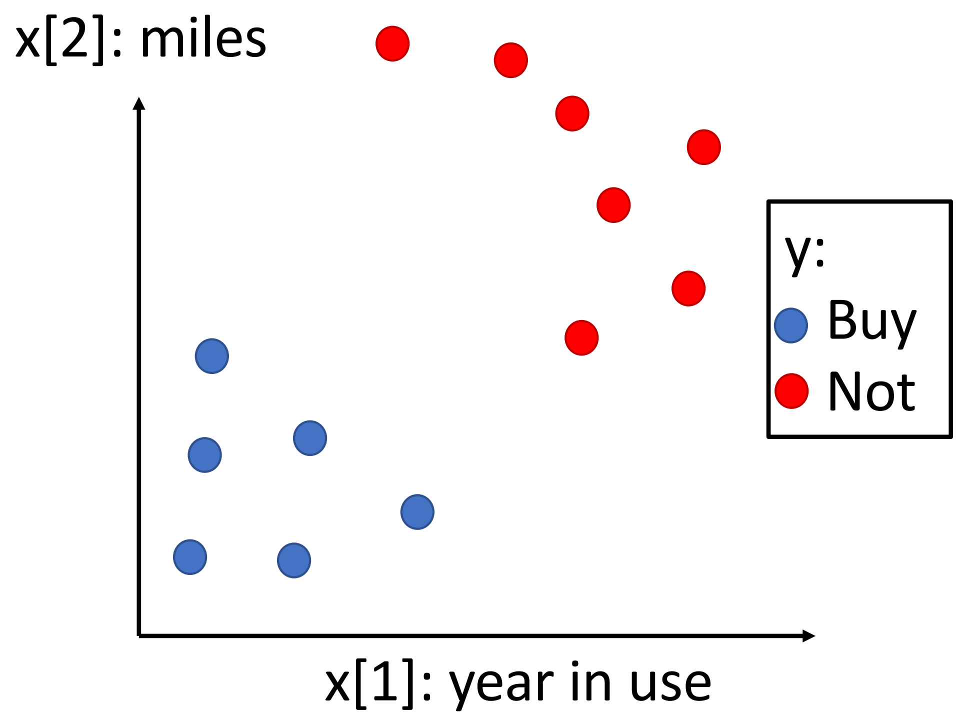
- Regression (curve Fitting)
  - Supervised learning
  - Form: Data point → Desired real number (e.g., price, chance, etc.)

Machine learning: capture patterns from training data that can be generalized to future data

Classification vs. regression: training data

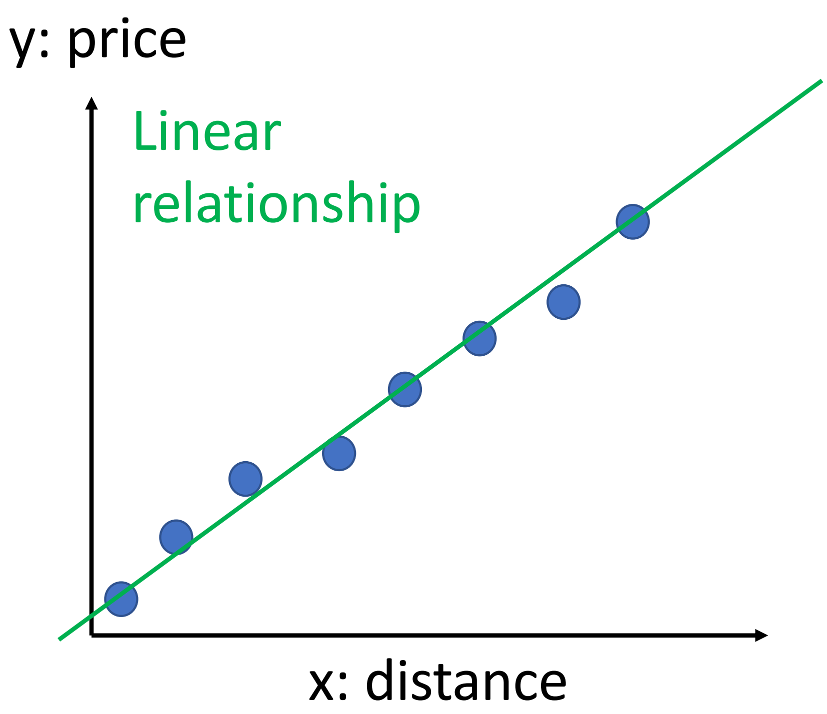
Regression (bus ticket):
From x (distance), predict y (price)

Classification (car buying company):
From ***x*** (year, miles), predict y (buy or not)

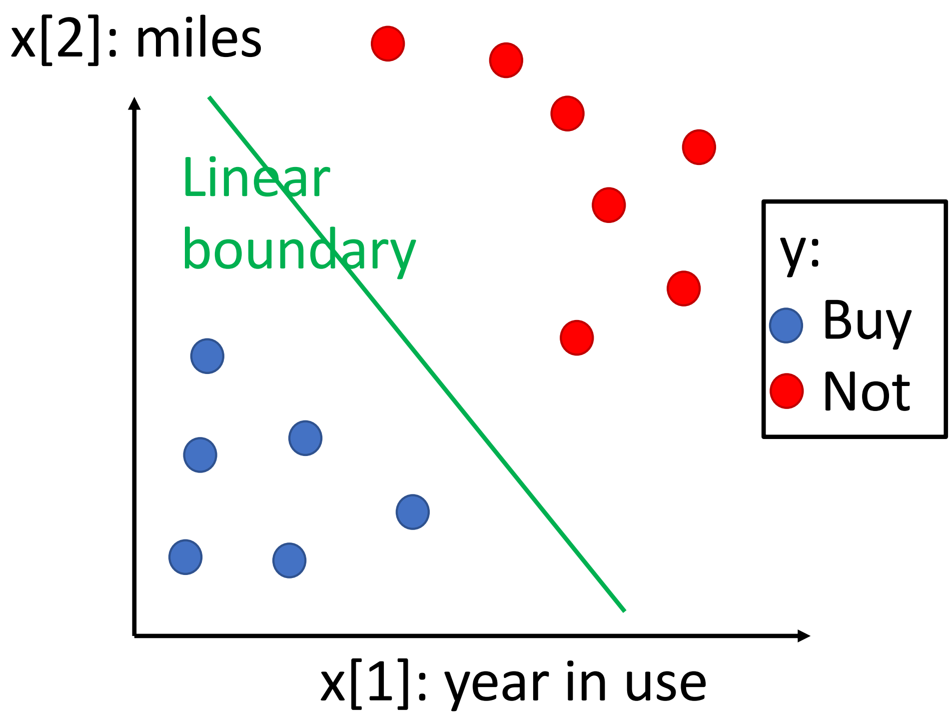


Classification vs. regression: find patterns

Regression (bus ticket):
From x (distance), predict y (price)

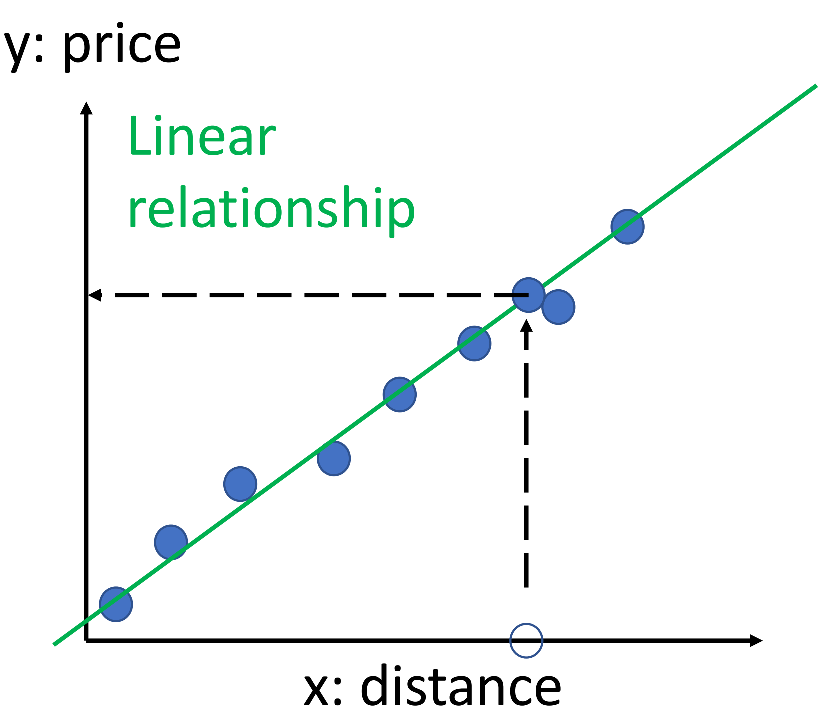


Classification (car buying company):
From ***x*** (year, miles), predict y (buy or not)

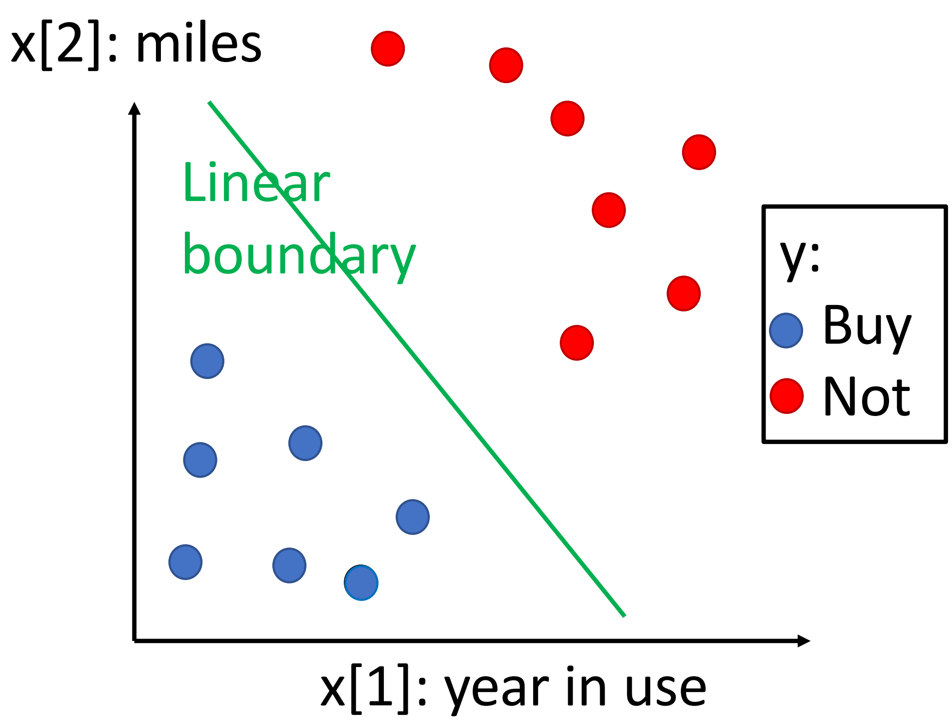


Classification vs. regression: generalization

Regression (bus ticket):
From x (distance), predict y (price)

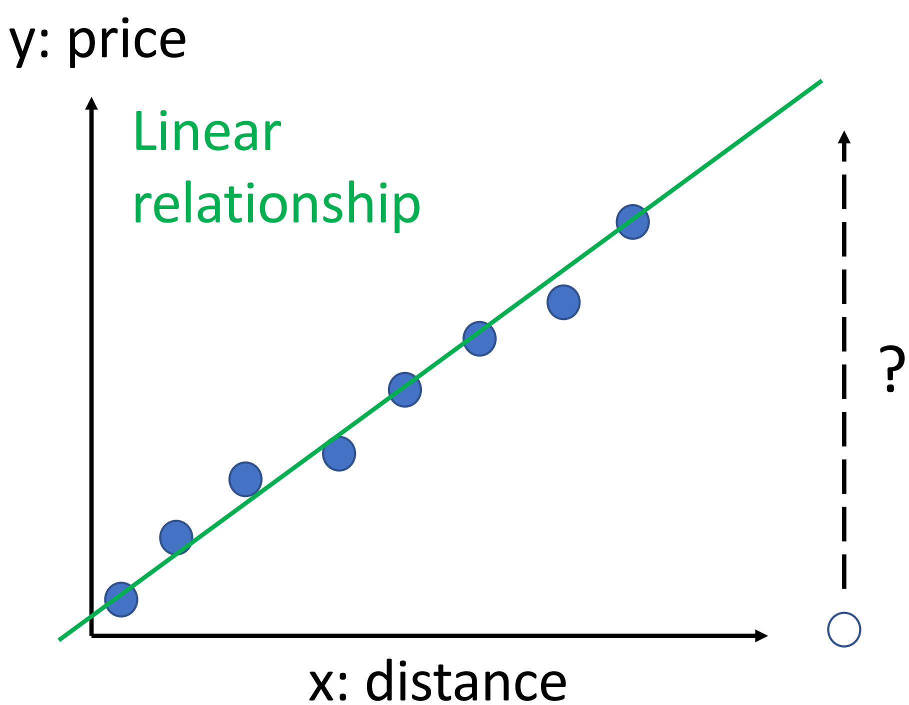


Classification (car buying company):
From ***x*** (year, miles), predict y (buy or not)

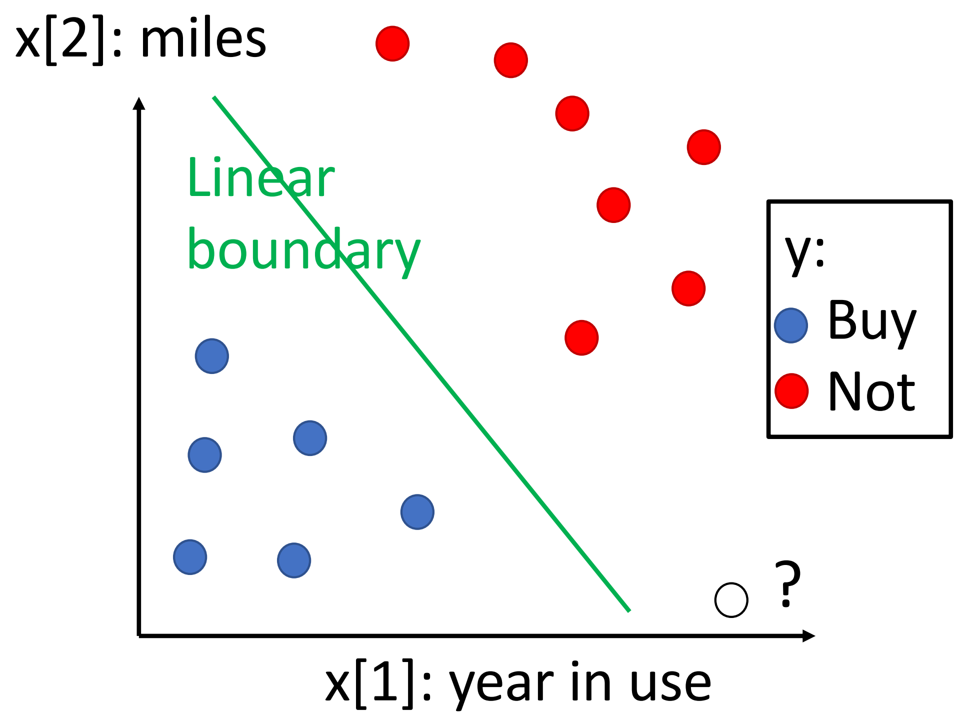


Classification vs. regression: generalization?

Regression (bus ticket):
From x (distance), predict y (price)



Classification (car buying company):
From ***x*** (year, miles), predict y (buy or not)



Supervised learning: PEAS

- Environment:
  - data distribution (e.g., pictures in the real world)

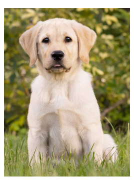
"Hat" here refers to the fact that this is a predicted label

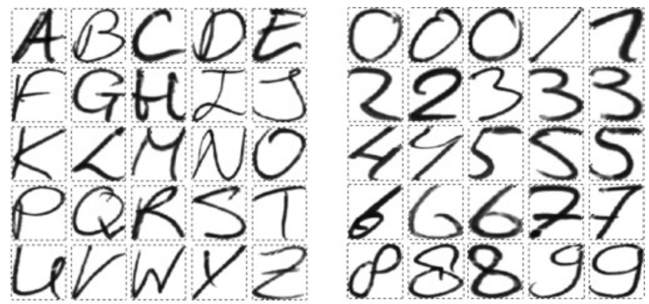
- Agent function (sensor, actuator):
  - input data: $\boldsymbol{x}$
  - the model to predict label: $\hat{y} = f(\boldsymbol{x})$ $\quad f: \mathcal{X} \mapsto \mathcal{Y}$

- Performance (loss/error) in comparison to the true label y:
  - Classification: $\mathbf{1}[y \neq f(\boldsymbol{x})]; \quad \mathbf{1}[\text{True}] = 1, \mathbf{1}[\text{False}] = 0$ ⟵ Indicator function
  - Regression: $\text{dis}\big(y, f(\boldsymbol{x})\big); \quad \text{e.g.}, \big(y - f(\boldsymbol{x})\big)^2 \text{ or } |y - f(\boldsymbol{x})|$

Supervised learning: training data

- Training data and test data are assumed to come from the same distribution
  - Train: real images of certain categories; Test: real images of the same categories (V)
  - Train: real images; Test: painting images (X)
  - Train: written digits; Test: written characters (X)





(a) (b)

  - More discussions in later lectures!

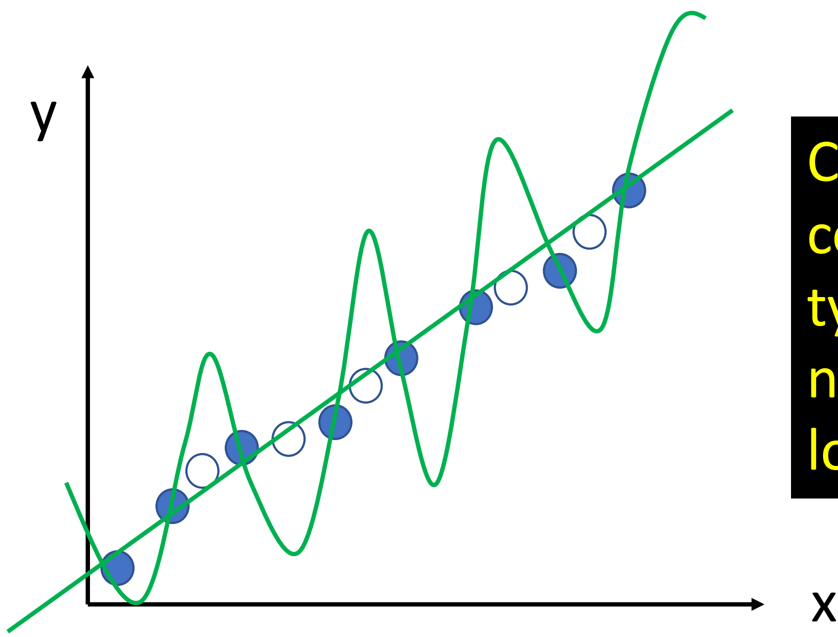
Machine learning: training vs. testing

- Patterns learned from the training data should be applicable for test data

$$D_{train} = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$$

$$D_{test} = \{(\boldsymbol{x}_1, y_{N+1}), \dots, (\boldsymbol{x}_N, y_{N+M})\}$$

In finding patterns, we only see training data!



Choosing a more complicated pattern type does not necessarily lead to lower test errors!

Machine learning: parameter estimation

- Pick a pattern/model type (e.g., linear curves or nonlinear curves)
- Pick an error function (e.g., SSE)
- Minimize the error on the test data

- How do we know if the learned model can be applicable to the test data?
  - Learning theory
  - Rule of thumb: more training data, more applicable!

- How to choose the model type?
  - Save some training data as "pseudo" test data!

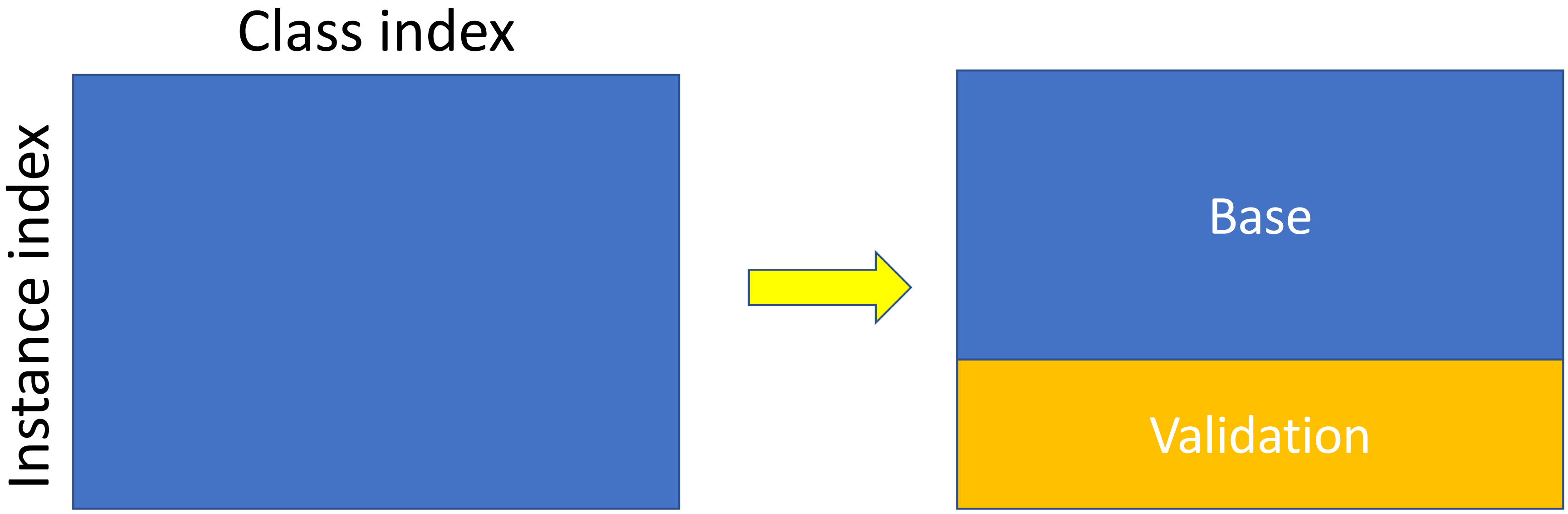
$$\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_N, y_N)\}$$

$$\Downarrow$$

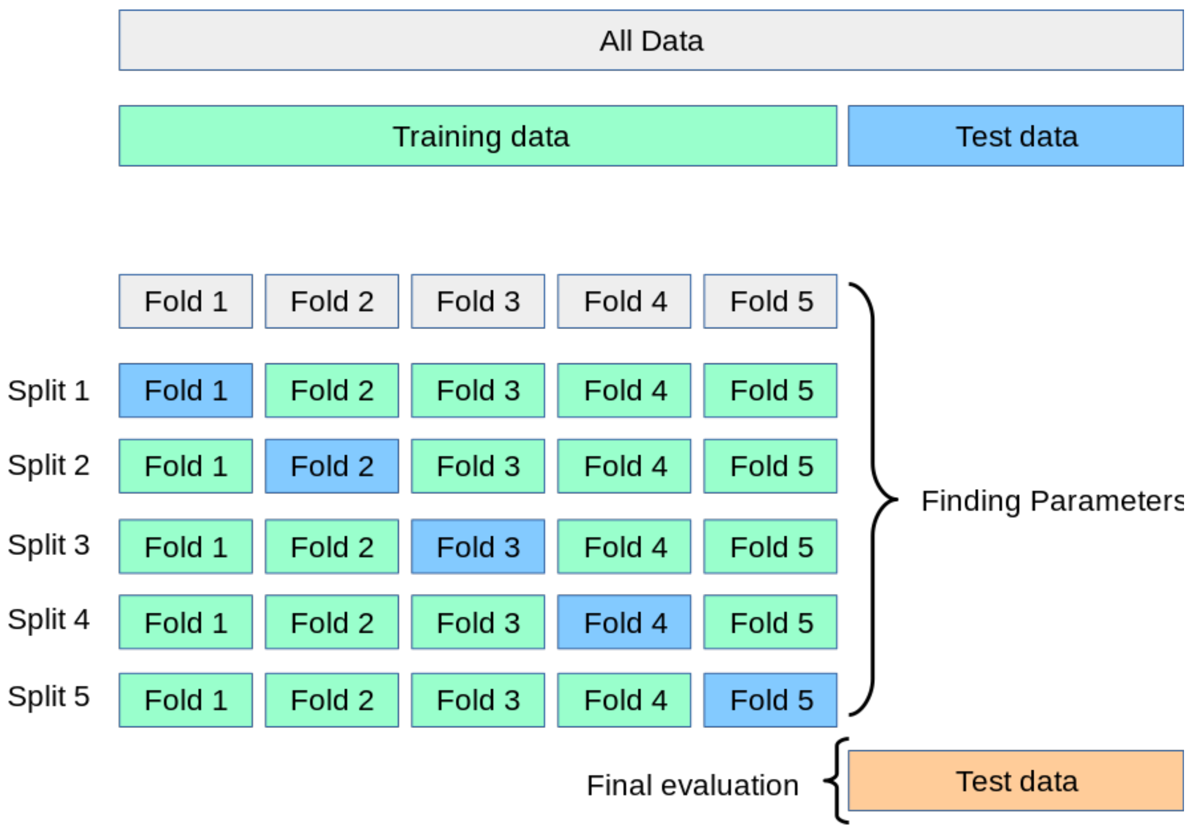
$$\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_{N-K}, y_{N-K})\}$$

$$\{(\boldsymbol{x}_{N-K+1}, y_{N-K+1}), \dots, (\boldsymbol{x}_N, y_N)\}$$

(Cross) validation



Cross validation



Supervised Classifier Examples

- Decision Tree
- Naïve Bayes
- Logistic Regression
- Support Vector Machine: SVM
- K-nearest neighbors: KNN
- Boosting/Bagging

$$\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$$

$$\downarrow$$

$$\boldsymbol{x}$$

$$\downarrow$$

$$y$$