The Supervised Learning Problem

Starting point:

- Outcome measurement Y (also called dependent variable, response, target)

- Vector of p predictor measurements X (also called inputs, regressors, covariates, features, independent variables)

- In the regression problem, Y is quantitative (e.g income, price, blood pressure).

- In the classification problem, Y takes values in a finite, unordered set (survived/died, digit 0-9, cancer class of tissue sample, spam/legit email).

- We have training data$(x_1, y_1), \ldots, (x_n, y_n)$ which are observations (examples, instances) of these measurements

Goal:

On the basis of the training data we want to

- Accurately predict unseen test cases

- Understand which inputs affect the outcome, and how they do so

- Assess the quality of our predictions and inferences

The Supervised Learning Problem

Starting point:

- Outcome measurement Y (also called dependent variable, response, target)
- Vector of p predictor measurements X (also called inputs, regressors, covariates, features, independent variables)
- In the regression problem, Y is quantitative (e.g income, price, blood pressure).
- In the classification problem, Y takes values in a finite, unordered set (survived/died, digit 0-9, cancer class of tissue sample, spam/legit email).
- We have training data$(x_1, y_1), \ldots, (x_n, y_n)$ which are observations (examples, instances) of these measurements

Goal:

On the basis of the training data we want to

- Accurately predict unseen test cases
- Understand which inputs affect the outcome, and how they do so
- Assess the quality of our predictions and inferences

Unsupervised Learning