

Ch1

Contents

A Statistical Learning

A.1	Some context of modern machine learning developments:
A.2	Types of Learning
A.3	Supervised Learning
A.4	Objective
A.5	Regression

Textbook: James et al. 1ed.

A Statistical Learning

[\[ToC\]](#)

A.1 Some context of modern machine learning developments:

- Brief History of AI
- Brief History of ML
- Deep Blue (Chess program)
- AlphaGo (Go program)

A.2 Types of Learning

- Supervised Learning
- Unsupervised Learning
- Deep Learning
- Reinforcement Learning
- Generative Adversarial Network (GAN)

A.3 Supervised Learning

- Input variables (Predictor, Feature, Covariates, Independent variable)
- Output variables (Response, Dependent variable)
- General relationship

$$Y = f(X) + \epsilon$$

A.4 Objective

$$Y = f(X) + \epsilon$$

- Want to estimate $f(\cdot)$
- Want to predict value of Y for next given value of X

A.5 Regression

- X = Total spending in advertisements
- Y = Total sales of next month

Name	Description
Auto	Gas mileage, horsepower, and other information for cars.
Boston	Housing values and other information about Boston suburbs.
Caravan	Information about individuals offered caravan insurance.
Carseats	Information about car seat sales in 400 stores.
College	Demographic characteristics, tuition, and more for USA colleges.
Default	Customer default records for a credit card company.
Hitters	Records and salaries for baseball players.
Khan	Gene expression measurements for four cancer types.
NCI60	Gene expression measurements for 64 cancer cell lines.
OJ	Sales information for Citrus Hill and Minute Maid orange juice.
Portfolio	Past values of financial assets, for use in portfolio allocation.
Smarket	Daily percentage returns for S&P 500 over a 5-year period.
USArrests	Crime statistics per 100,000 residents in 50 states of USA.
Wage	Income survey data for males in central Atlantic region of USA.
Weekly	1,089 weekly stock market returns for 21 years.

(TABLE 1.1. from ISLR.)

Dataset used in the book

```
# install.packages("ISLR") # if you haven't installed the package
```

```
library(ISLR)      # load ISLR package  
data(Advertising) # Advertising not in ISLR. Must go to author's website
```

```
Adv <- read.csv(file="http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv", header=T)  
names(Adv)  # see column names  
head(Adv)   # just see first couple of lines  
dim(Adv)    # see the size of data
```

```
Adv <- Adv[, -1] # remove the first column "X"
```

```
pairs(Adv)
```

```
Inc <- read.csv(file="http://www-bcf.usc.edu/~gareth/ISL/Income2.csv", header=T)  
names(Inc)  # see column names  
head(Inc)   # just see first couple of lines  
dim(Inc)    # see the size of data
```

```
Inc <- Inc[, -1] # remove the first column "X"
```

```
pairs(Inc)
```