Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Contents

Outline

1. Linear Models
   - Framework and notations
   - Toy example : Housing price prediction
   - Toy example (from the book) : Sales prediction
   - Linear model properties
   - Fitting the regression
   - Solving the regression analytically
   - Gradient descent principles
   - Other algorithms LDA, Polynomial expansion
   - Summary

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). X is a matrix of dimension (N,p), where n is the number of measurements ;
- In the **regression problem**, Y is quantitative (e.g price, blood pressure) ;
- We have training data (x1,y1),..., (xN,yN). These are observations (examples, instances) of these measurements.

Outline

Linear Regression model with one variable

$Y_i = \beta_0 + \beta_1 X_1$

$Price_{house30} = K + \beta_1 Surface_{House30}$

Figure (TBD) In fact, we could imagine the price depends from severals factors, so we come with Linear Regression with several variables :

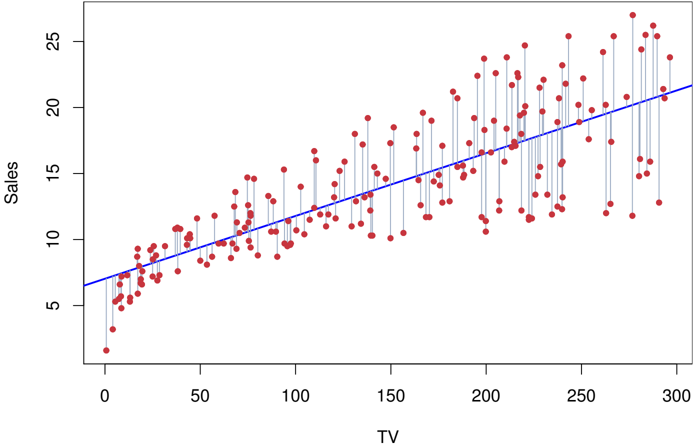
$Price_{house30} =$
$K + \beta_1 Surface_{House30} + \beta_2 NbOfRooms_{House30} + \beta_3 Location_{House30}$

In general :

$Y_i = h(X^i) = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + ... + \beta_i X_{i,p}$

Traditionaly p is called the number of features. We will use matrix notation, so there will be double indices.

Outline



$$Sales \approx \beta_0 + \beta_1 TV$$

Outline

- The parameters in the linear regression model are very easy to interpret.
- $\beta_j, 1 \leq j \leq p$ is the average increase in Y when X_j is increased by one and all other X_i are held constant.
- Vocabulary : β_0 is the intercept (i.e. the average value for Y if all the X ?s are zero), β_j is the slope for the jth variable X_j

Outline

- Historical method : least square regression ;
- Modern method : numerical iterative process : gradient descent and a huge family of similar algorithms (Maths : (Numerical)(Convex or not) Optimization.

Cost function, traditionally noted $J(\beta)$ is given by :
$J(\beta) = \frac{1}{2n} \sum_{i=1}^{n} (h(X^i) - Y_i)^2$
n is N explained before.

Outline

Explanation,
Solution in One dim
Solution in p dim

-

Outline

1. Linear Models
   - Framework and notations
   - Toy example : Housing price prediction
   - Toy example (from the book) : Sales prediction
   - Linear model properties
   - Fitting the regression
   - Solving the regression analytically
   - **Gradient descent principles**
   - Other algorithms LDA, Polynomial expansion
   - Summary

- Level on a curve or surface, direction of steppest descent
- Stochastic approach (SGD)

Outline

- The goal of Unsupervised Learning is to discover interesting things about the measurements : is there an informative way to visualize the data ? Can we discover subgroups among the variables or among the observations ?
- We discuss two methods :
  - **principal components analysis (PCA)**, a tool used for data visualization or data pre-processing before supervised techniques are applied, and
  - **clustering**, a broad class of methods for discovering unknown subgroups in data.

Outline

Linearity is a limitation but solving principles are more general