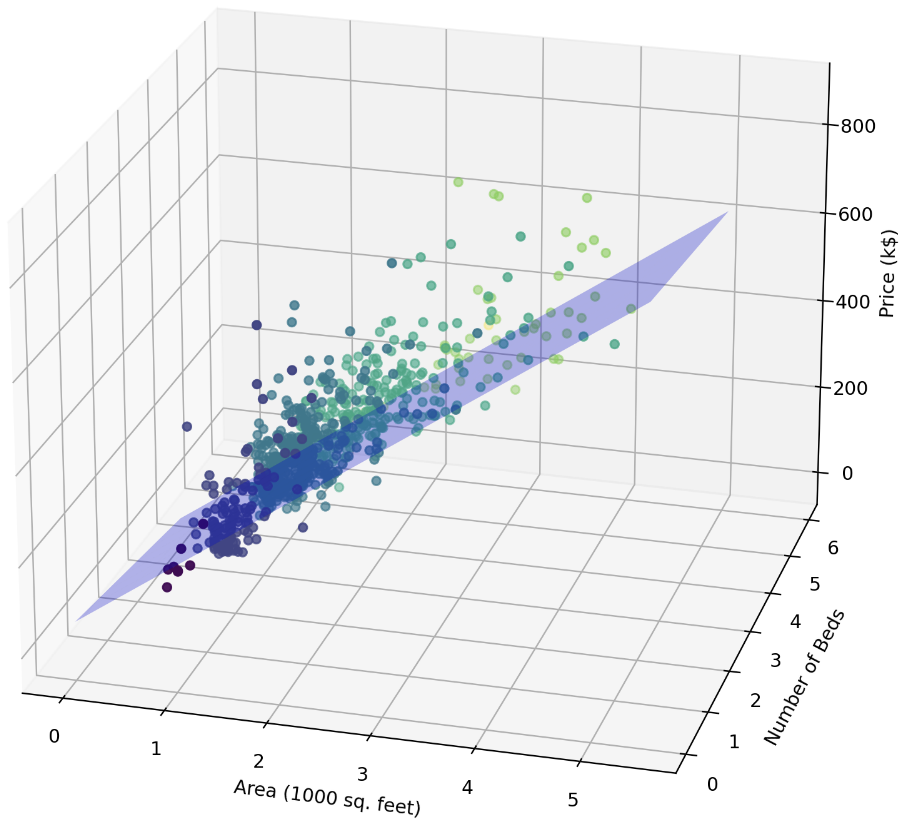
12 Least Squares Data Fitting



Unit 1: Vectors, Book ILA Ch. 1-5

Unit 2: Matrices, Book ILA Ch. 6-11 + Book IMC Ch. 2

Unit 3: Least Squares, Book ILA Ch. 12-14

- 11 Least Squares
- ***12 Least Squares Data Fitting***
- 13 Least Squares Classification

Outline: 12 Least Squares Data Fitting

- Least Square Model Fitting
- Validation
- Feature Engineering

True relationship: f

Definition: When we believe that a scalar y and an n-vector x are related by model:

$$y \approx f(x),$$

we use the following vocabulary:

- x is called the independent variable
- y is called the outcome or response variable
- $f : \mathbb{R}^n \to \mathbb{R}$ represents the "true" relationship between x and y.

Generally, we do not know f, we just assume it exists. Our goal is to learn f, or a reasonable approximation of it, using data.

Data

$\color{brown}{\text{Definition}}$: The data:

$$x^{(1)}, \ldots, x^{(N)}, y^{(1)}, \ldots, y^{(N)}$$

are called observations, examples, samples, or measurements.

- $x^{(i)}, y^{(i)}$ is ith data pair
- $x_j^{(i)}$ is the jth component of ith data point $x^{(i)}$.

Model: $\hat{f}$

$\color{brown}{\text{Definition}}$: Choosing a set of basis functions: $f_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1 \ldots p$, we model a guess or approximation of f as:

$$\hat{f}\,(x) = \theta_1 f_1(x) + \ldots + \theta_p f_p(x),$$

where:

- θ_i are model parameters that we will learn from the data,
- $\hat{y}^{(i)} = \hat{f}\,(x^{(i)})$ is (the model's) prediction of $y^{(i)}$.

$\color{brown}{\text{Remark}}$: If our model is good, then $\hat{y}^{(i)} \approx y^{(i)}$, i.e., model is consistent with observed data.

Residuals

$\color{brown}{\text{Definition}}$: Given:

- observations $x^{(1)}, \ldots, x^{(N)}, \ldots, y^{(1)}, y^{(N)}$,
- a model $\hat{f}$ generating $\hat{y}^{(i)} = \hat{f}\,(x^{(i)})$ predictions of $y^{(i)}$, for $i = 1, \ldots, p$,

we define the prediction error, or residual:

$$r_i = y^{(i)} - \hat{y}^{(i)}.$$

Least Square Data Fitting

Definition: The Least Square Data Fitting problem is the problem of choosing model's parameters $\theta_1, \ldots, \theta_n$ that minimize the RMS prediction error on the dataset:

$$\left(\frac{r_1^2 + \ldots + r_N^2}{N}\right)^{1/2}.$$

LS Data Fitting and LS

The Least Square (LS) Data Fitting problem can be formulated as a Least Squares (LS) Problem.

Notations: We can express $y^{(i)}, \hat{y}^{(i)}$, and r_i as N-vectors:

- $y = (y^{(1)}, \ldots, y^{(N)})$ is vector of outcomes,
- $\hat{y} = (\hat{y}^{(1)}, \ldots, \hat{y}^{(N)})$ is vector of predictions,
- $r = (r_1, \ldots, r_N)$ is vector of residuals.

Proposition: Define the $N \times p$ matrix A with elements $A_{ij} = f_j(x^{(i)})$, such that $\hat{y} = A\theta$. The least square data fitting problem amounts to choose θ that minimizes:

$$||A\theta - y||^2,$$

which shows that it can be written as a Least Square Problem.

Solving the LS Data Fitting Problem

Proposition: Consider a LS Data Fitting problem formulated as minimizing $||A\theta - y||^2$. Assuming that the columns of A are independent, the solution is:

$$\hat{\theta} = (A^T A)^{-1} A^T y.$$