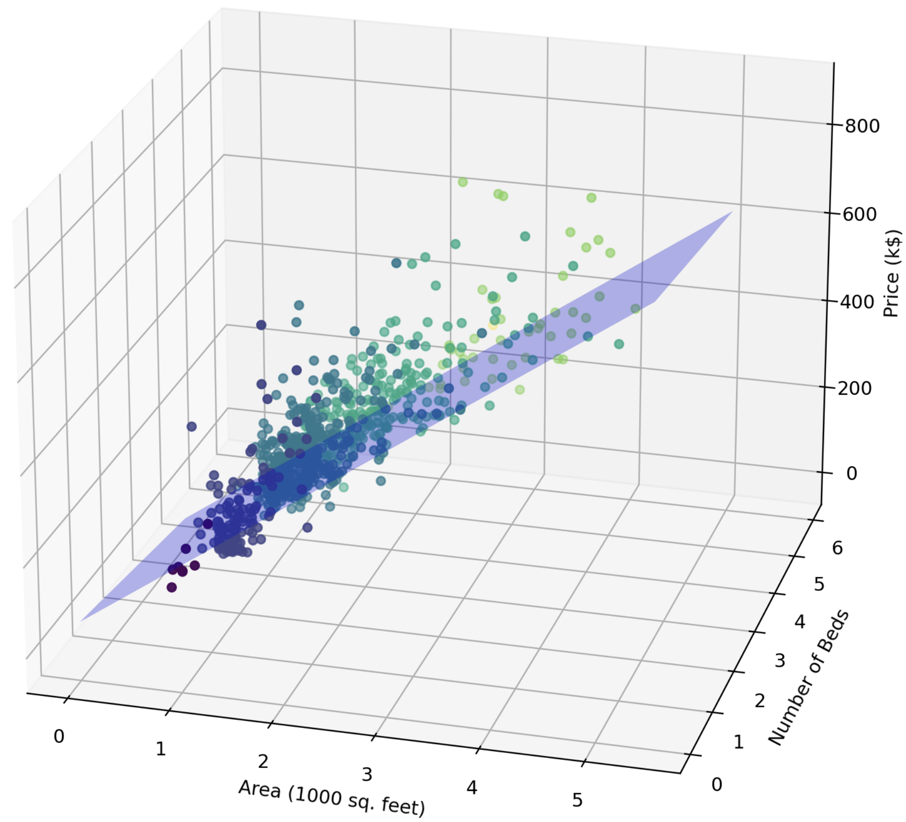
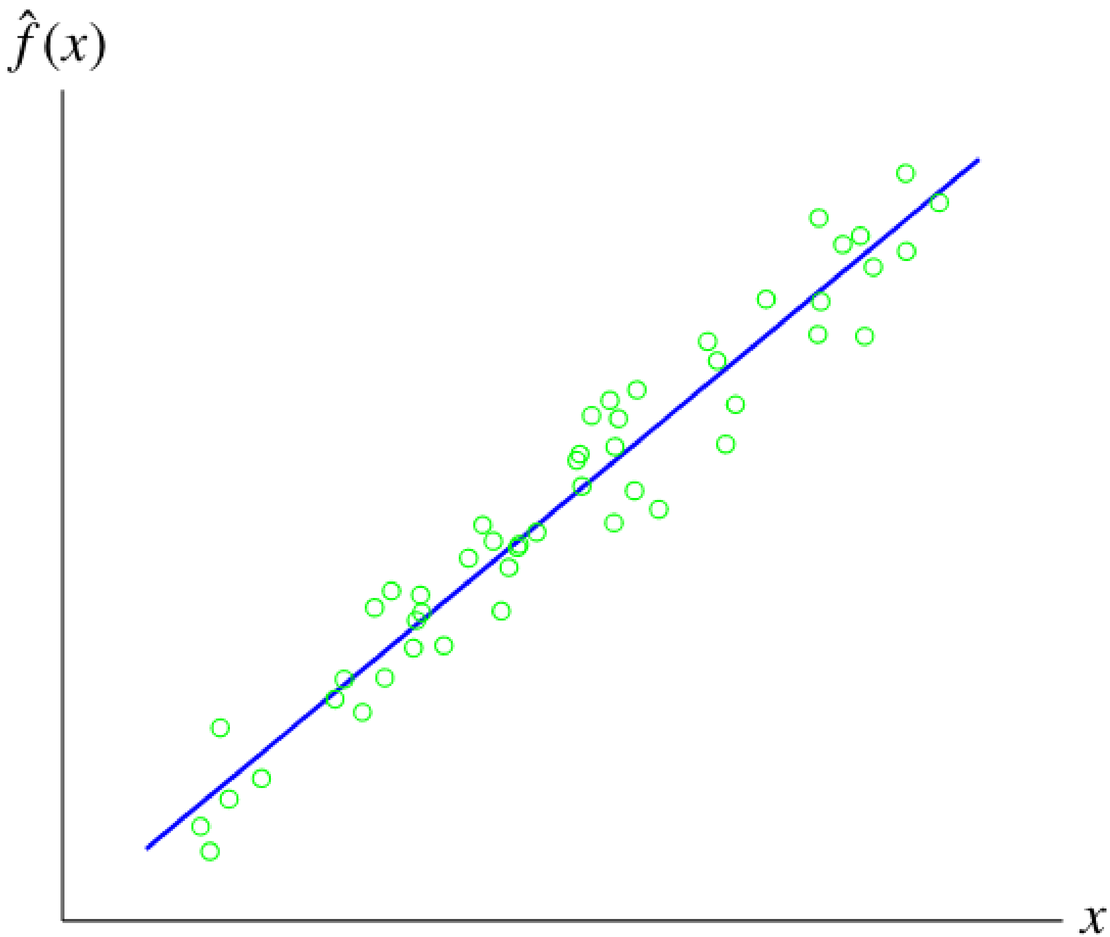
12 Least Squares Data Fitting





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: Regression and Classification***

Outline: 12 Least Squares Data Fitting

- Least Square Model Fitting
- Application to Regression and Classification

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.

Model: $\hat{f}$

Definition: Choosing a set of basis functions: $f_j : \mathbb{R}^n \to \mathbb{R}$, for $j = 1...p$, 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_j are parameters that we learn from data: $x^{(1)}, \ldots, x^{(N)}, \ldots, y^{(1)}, \ldots y^{(N)}$,
- $\hat{y}^{(i)} = \hat{f}(x^{(i)})$ is (the model's) prediction of $y^{(i)}$, for $i = 1, \ldots, N$.

Remark: If our model is good, then $\hat{y}^{(i)} \approx y^{(i)}$.

Least Square Data Fitting

Definition: We define the prediction error, or residual for each $i = 1, \ldots, N$:

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

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:

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

Proposition: Define the $N \times p$ matrix A with elements $A_{ij} = f_j(x^{(i)})$, such that $\hat{y} = A\theta$, where $y = (y^{(1)}, \ldots, y^{(N)})$ is vector of outcomes. 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. Assuming that the columns of A are independent, the solution is:

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

Outline: 12 Least Squares Data Fitting

Classification and Regression

Artificial Intelligence (AI): Techniques that enable machines to mimic human intelligence.

Machine Learning (ML): Techniques that enable machines to learn from data.

Supervised Learning: Task of learning a function that maps an input to an output based on example input-output pairs.
Examples:
- **Regression:** maps an input to a quantitative value.
- **Classification:** maps an input to a categorical value.

Unsupervised Learning: Task of discovering any naturally occurring patterns in a data set.
Examples:
- **Clustering:** discover groups (clusters) within the data: today.
- **Dimension reduction:** later in this class.

Regression of House Prices

- x: house's area in 1000 sq feet,
- y: house's price in in k$ (we do not consider the number of beds for simplicity).

Consider the model: $\hat{f}(x) = \theta_1 f_1(x) + \theta_2 f_2(x)$ with $f_1(x) = 1$ and $f_2(x) = x$, i.e.:

$$\hat{f}(x) = \theta_1 + \theta_2 x.$$

Example: What are A, y? Explain how you can find $\hat{\theta}_1$ and $\hat{\theta}_2$ with Python. At home: compute $\hat{\theta}_1$ and $\hat{\theta}_2$ manually (hard).

From Regression to (Binary) Classification

- Model $\hat{f}$ outputs a number.
- Binary classification wants a category: +1 or -1 only.

$\rightarrow$ use $sign(\hat{f})$ in place of $\hat{f}$ to classify.

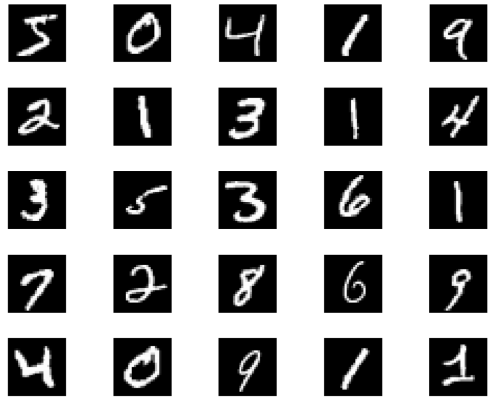
Classification of MNIST

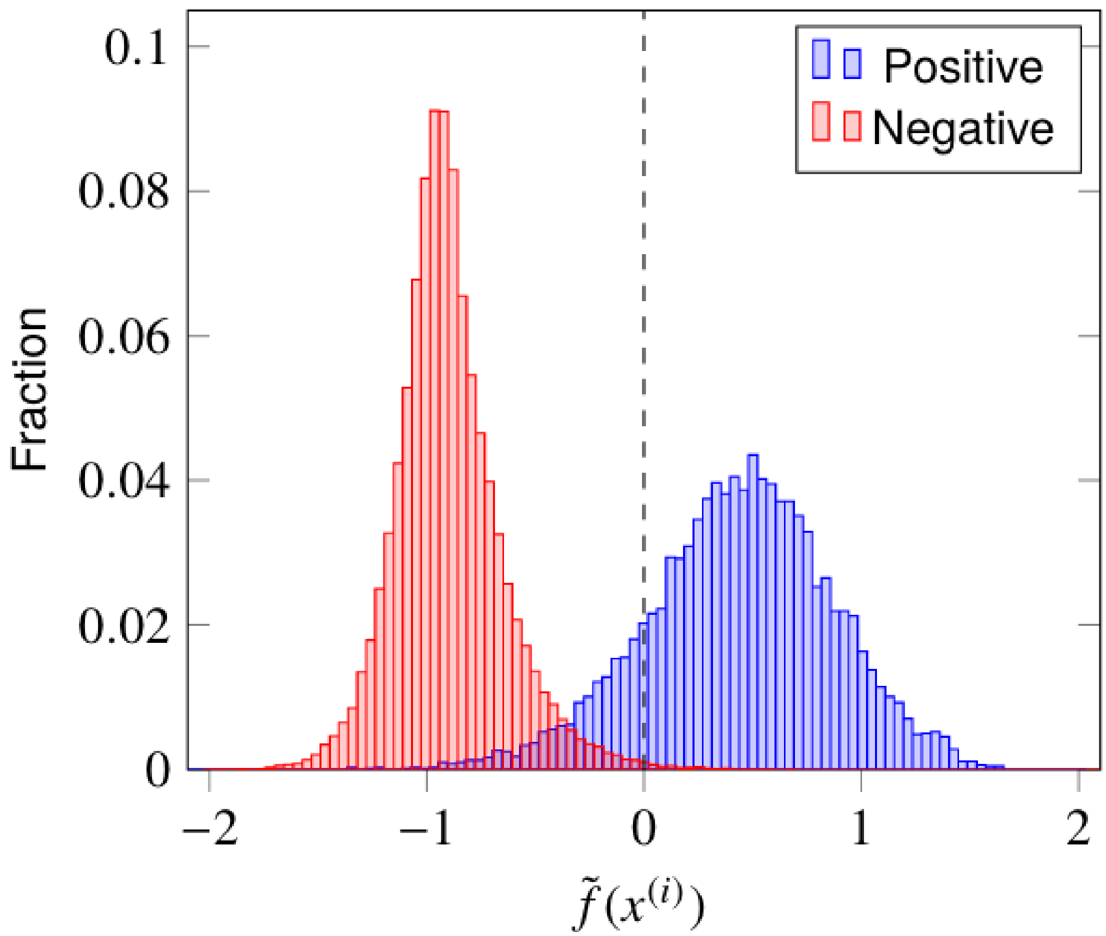
- x image of size 28×28 from the MNIST dataset,
- y: whether the image shows a 0 digit or another digit.

Consider the model: $\hat{f}(x) = sign(\theta_0 f_0(x) + \theta_1 f_1(x) + \ldots + \theta_{784} f_{784}(x))$ with $f_0(x) = 1$, and $f_p(x) = x_p$ for $p = 1, \ldots, 784$, i.e.:

$$\hat{f}(x) = sign(\theta_0 + \theta_1 x_1 + \ldots + \theta_{784} x_{784}).$$

Example: What are x, y, A? Explain how you can find the $\hat{\theta}$s with Python.





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: Regression and Classification***

Resources: Book ILA Ch. 13-14.