Linear regression

Supervised learning example

$a(x)$ → 13

Supervised learning

x_i — example

y_i — target value

$x_i = (x_{i1}, \ldots, x_{id})$ — features

$X = \big((x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\big)$ — training set

$a(x)$ — model, hypothesis

$$x \longrightarrow a(x) \longrightarrow y^{pred}$$

Regression and classification

$y_i \in \mathbb{R}$ — regression task

- Salary prediction
- Movie rating prediction

Regression and classification

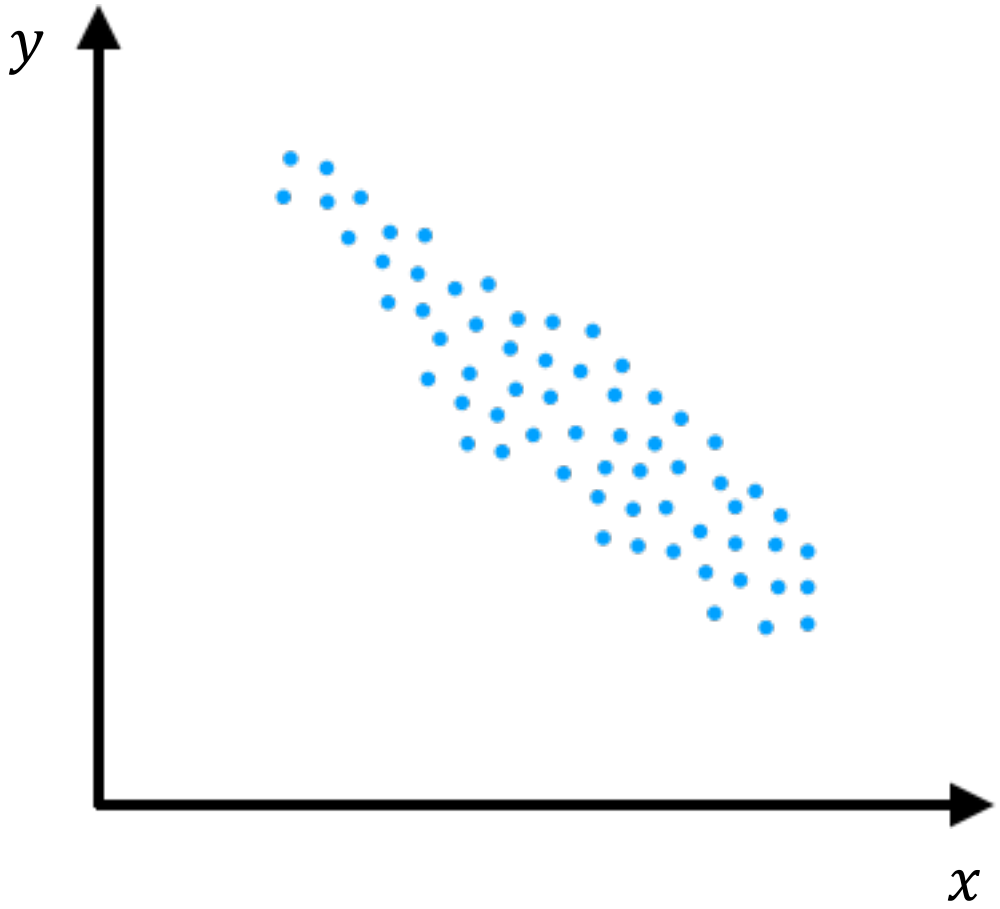
$y_i \in \mathbb{R}$ — regression task

- Salary prediction
- Movie rating prediction

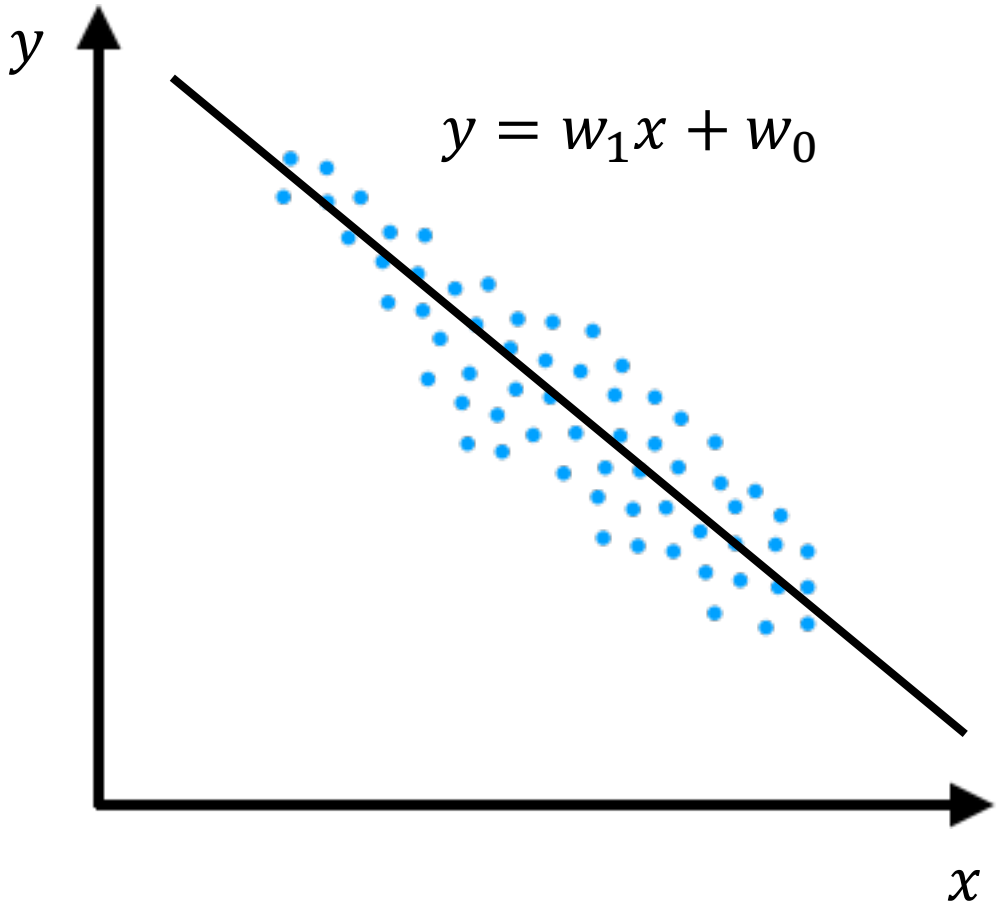
y_i belongs to a finite set — classification task

- Object recognition
- Topic classification

Linear model for regression example



Linear model for regression example



Linear model for regression

$$a(x) = b + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$$

- $w_1, \ldots, w_d$ — coefficients (weights)
- b — bias
- $d + 1$ parameters
- To make it simple: there's always a constant feature

Linear model for regression

Vector notation:

$$a(x) = w^T x$$

For a sample X:

$$a(X) = Xw$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{\ell 1} & \cdots & x_{\ell d} \end{pmatrix}$$

Loss function

How to measure model quality?

Mean squared error:

$$L(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}(w^T x_i - y_i)^2$$

$$= \frac{1}{\ell}\|Xw - y\|^2$$

Training a model

Fitting a model to training data:

$$L(w) = \frac{1}{\ell} \|Xw - y\|^2 \to \min_w$$

Training a model

Fitting a model to training data:

$$L(w) = \frac{1}{\ell} \|Xw - y\|^2 \to \min_{w,}$$

Exact solution:

$$w = (X^T X)^{-1} X^T y$$

But inverting a matrix is hard for high-dimensional data!

Summary

- Linear models are very simple
- MSE can be used as a loss function
- There is an analytical solution, but we need more generic and scalable learning method