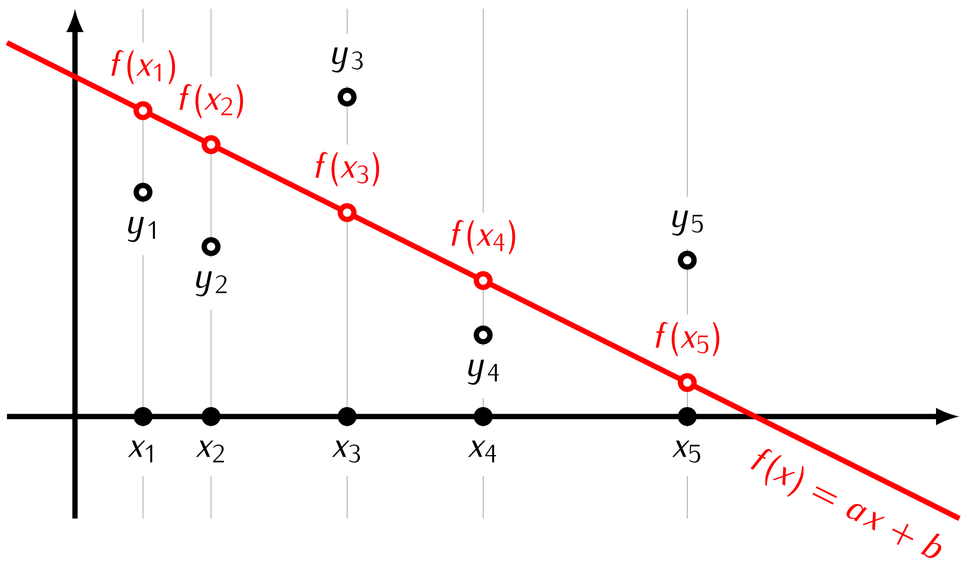
L_1 regression

Problem. Given points with coordinates (x_i, y_i) for $i = 1, \ldots, n$ find a function $f(x) = ax + b$ such that the sum

$$\sum_{i=1}^{n} |f(x_i) - y_i|$$

is a as small as possible.



Note. Compare with L_2 regression (least squares): we want to minimize

$$\sqrt{\sum_{i=1}^{n} |f(x_i) - y_i|^2}$$

Problem. Given points with coordinates (x_i, y_i) for $i = 1, \ldots, n$ find a function $f(x) = ax + b$ such that the sum

$$\sum_{i=1}^{n} |f(x_i) - y_i| = \sum_{i=1}^{n} |ax_i + b_i - y_i|$$

is a as small as possible.

<u>Decision variables:</u> $a, b \in \mathbb{R}$

We want to minimize

$$\sum_{i=1}^{n} |ax_i + b_i - y_i|$$

<u>Problem:</u> What to do with the absolute value?

<u>Solution:</u> Introduce new variables $e_1, \ldots, e_n$ such that

$$e_i \geqslant |ax_i + b - y_i|$$

then minimize $e_1 + e_2 + \ldots + e_n$.

<u>Note:</u> $e_i \geqslant |ax_i + b_i|$ if and only if the following hold:

1) $e_i \geqslant 0$
2) $e_i \geqslant ax_i + b - y_i$
3) $e_i \geqslant -(ax_i + b - y_i) = -ax_i - b + y_i$

<u>Linear program:</u>

Decision variables: $a, b, e_1, \ldots, e_n$

We want to minimize

$$z = e_1 + e_2 + \ldots + e_n$$

<u>Constraints:</u>

$e_i \geqslant ax_i + b - y_i$ (or: $y_i \geqslant ax_i + b - e_i$)

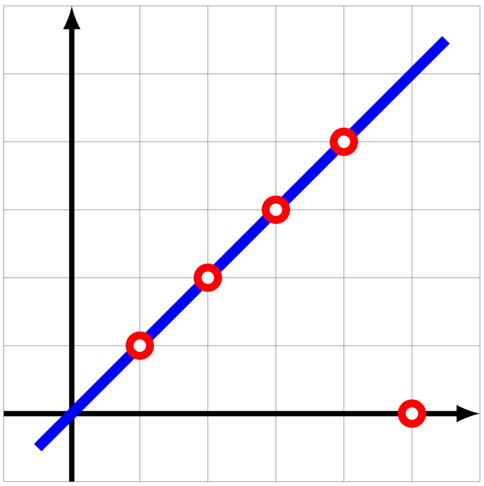
$e_i \geqslant -ax_i - b + y_i$ (or: $-y_i \geqslant -ax_i - b - e_i$)

$e_1, \ldots, e_n \geqslant 0$

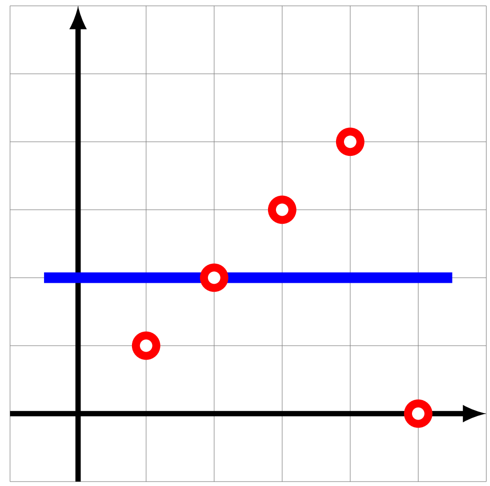
$a, b \in \mathbb{R}$

L_1 regression vs L_2 regression

- L_1 regression is less sensitive that L_2 if we change the value of a single point.



L_1 regression



L_2 regression

- L_2 regression gives a uniquely defined line if there are at least two points with different x-coordinates. L_1 regression can have infinitely many solutions.