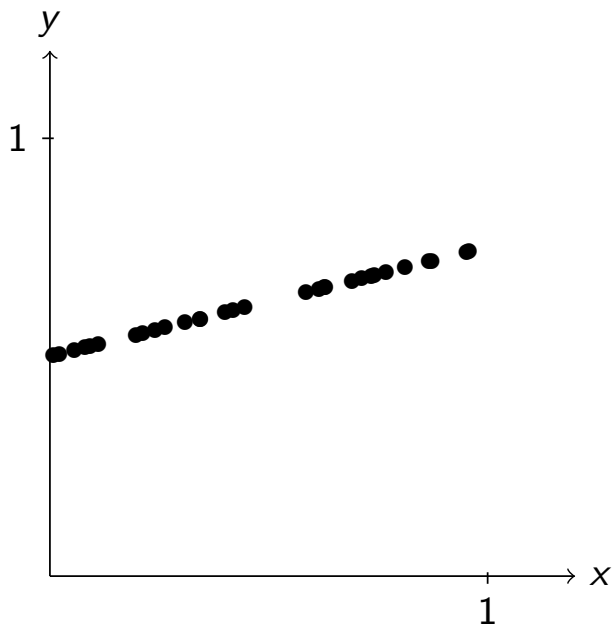
Parameter tuning

Line approximation

We have a few sample points and we want to find a line that approximate these points.



Our model is just a line

$$y_{pred} = ax + b$$

We want to find a and b to best match our samples.

First we have a define a loss that measures how good our predictions are.

$$l(x, y, a, b) = (y - y_{pred})^2 = (y - (ax + b))^2$$

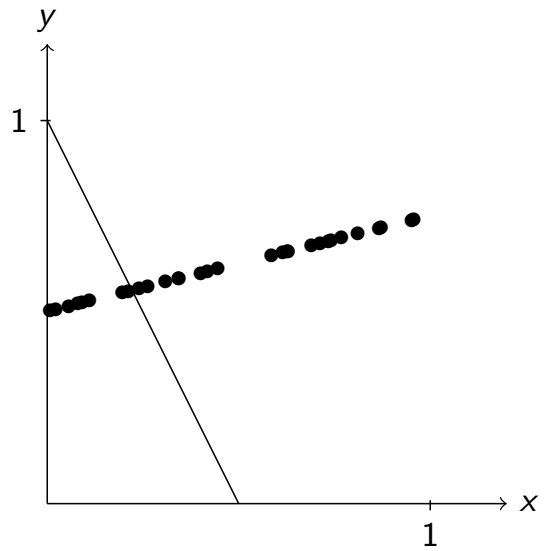
and now, we compute how the loss is affected by small changes of a and b:

$$\frac{dl}{da} = 2x(ax + b - y) \qquad \frac{dl}{db} = 2(ax + b - y)$$

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

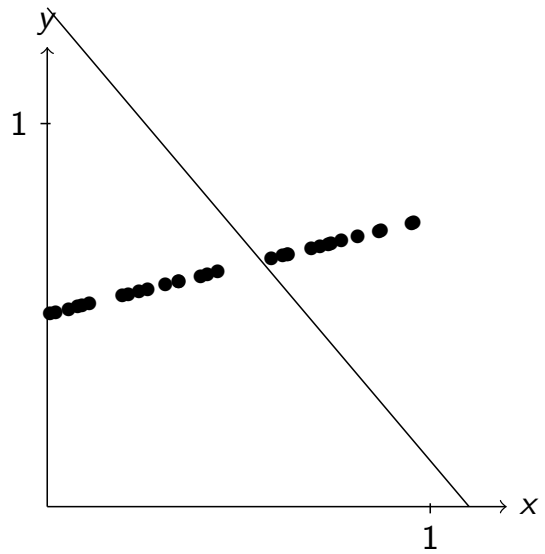
$$a = -2.00 \quad b = 1.00 \quad \overline{\frac{dl}{da}} = -1.07 \quad \overline{\frac{dl}{db}} = -1.37 \quad \overline{l(x, y, a, b)} = 0.860367$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

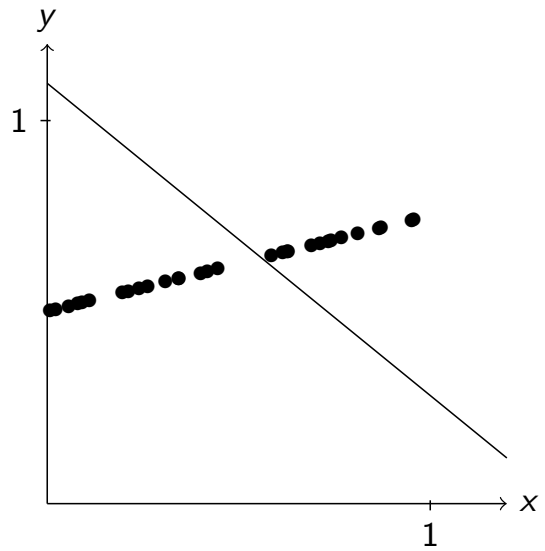
$$a = -1.18 \quad b = 1.30 \quad \overline{\frac{dl}{da}} = -0.17 \quad \overline{\frac{dl}{db}} = 0.09 \quad \overline{l(x, y, a, b)} = 0.159420$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

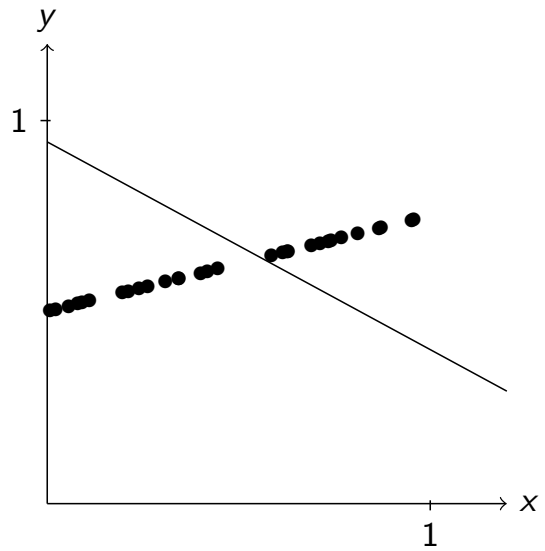
$$a = -0.82 \quad b = 1.10 \quad \overline{\frac{dl}{da}} = -0.13 \quad \overline{\frac{dl}{db}} = 0.07 \quad \overline{l(x, y, a, b)} = 0.088204$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

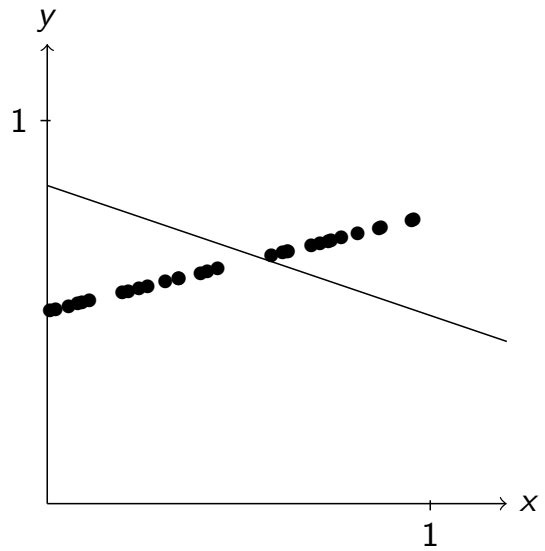
$$a = -0.54 \quad b = 0.94 \quad \overline{\frac{dl}{da}} = -0.09 \quad \overline{\frac{dl}{db}} = 0.05 \quad \overline{l(x, y, a, b)} = 0.048802$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

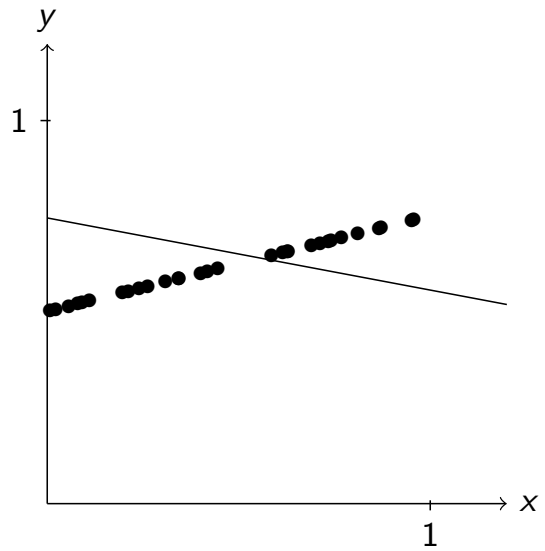
$$a = -0.34 \quad b = 0.83 \quad \overline{\frac{dl}{da}} = -0.07 \quad \overline{\frac{dl}{db}} = 0.04 \quad \overline{l(x, y, a, b)} = 0.027001$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

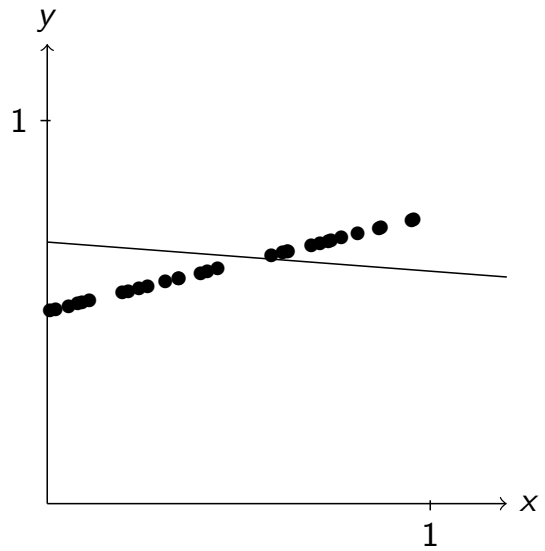
$$a = -0.19 \quad b = 0.75 \quad \overline{\frac{dl}{da}} = -0.05 \quad \overline{\frac{dl}{db}} = 0.03 \quad \overline{l(x, y, a, b)} = 0.014939$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

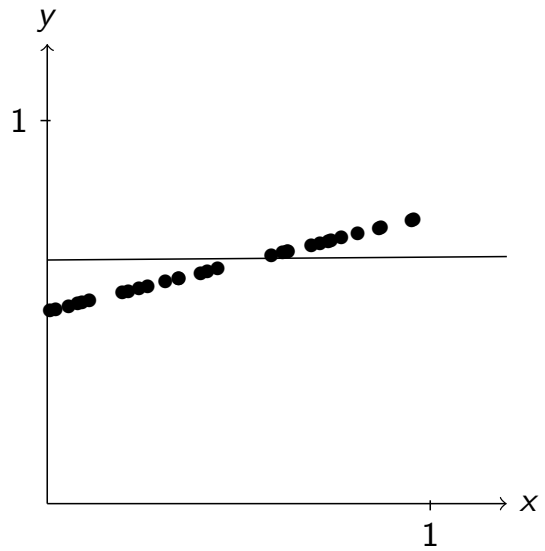
$$a = -0.08 \quad b = 0.68 \quad \overline{\frac{dl}{da}} = -0.04 \quad \overline{\frac{dl}{db}} = 0.02 \quad \overline{l(x, y, a, b)} = 0.008266$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

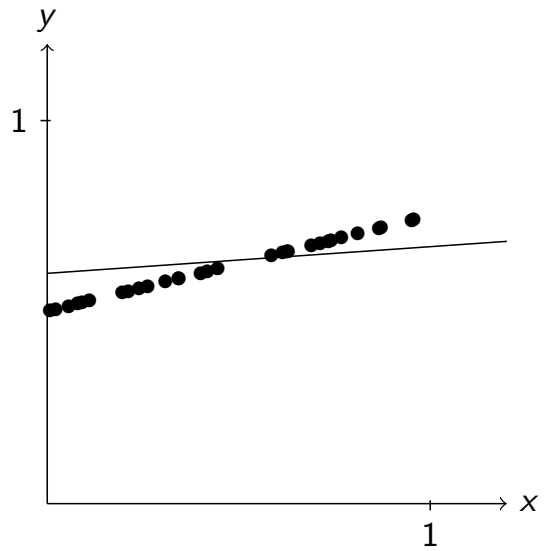
$$a = 0.01 \quad b = 0.64 \quad \overline{\frac{dl}{da}} = -0.03 \quad \overline{\frac{dl}{db}} = 0.02 \quad \overline{l(x, y, a, b)} = 0.004573$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

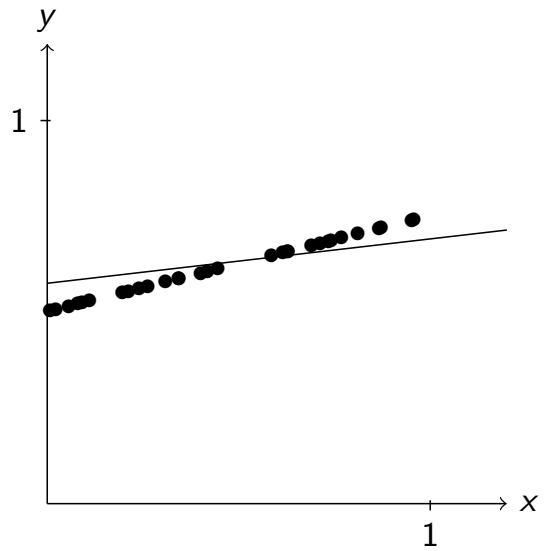
$$a = 0.07 \quad b = 0.60 \quad \overline{\frac{dl}{da}} = -0.02 \quad \overline{\frac{dl}{db}} = 0.01 \quad \overline{l(x, y, a, b)} = 0.002530$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

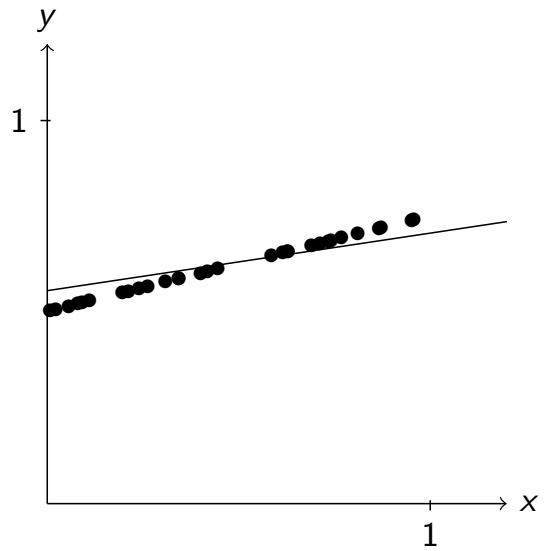
$$a = 0.12 \quad b = 0.58 \quad \overline{\frac{dl}{da}} = -0.02 \quad \overline{\frac{dl}{db}} = 0.01 \quad \overline{l(x, y, a, b)} = 0.001400$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

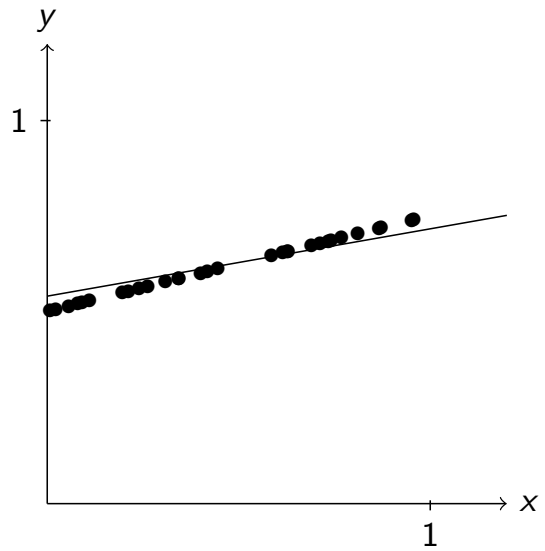
$$a = 0.15 \quad b = 0.56 \quad \overline{\frac{dl}{da}} = -0.01 \quad \overline{\frac{dl}{db}} = 0.01 \quad \overline{l(x, y, a, b)} = 0.000775$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

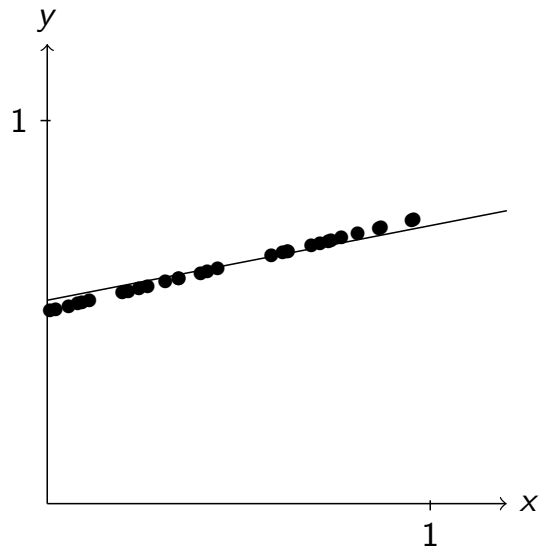
$$a = 0.18 \quad b = 0.54 \quad \overline{\frac{dl}{da}} = -0.01 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000429$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

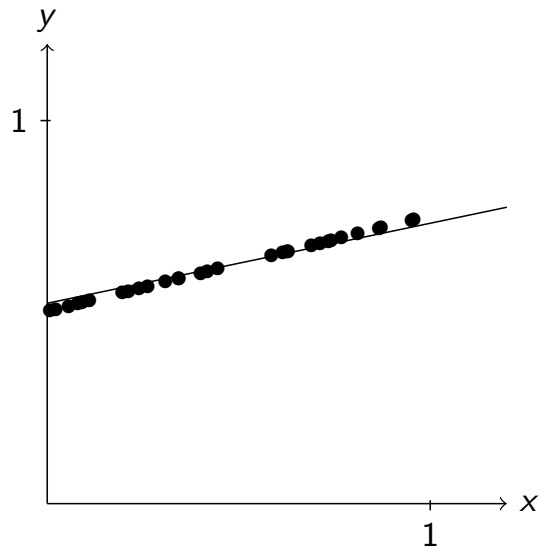
$$a = 0.19 \quad b = 0.53 \quad \overline{\frac{dl}{da}} = -0.01 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000237$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

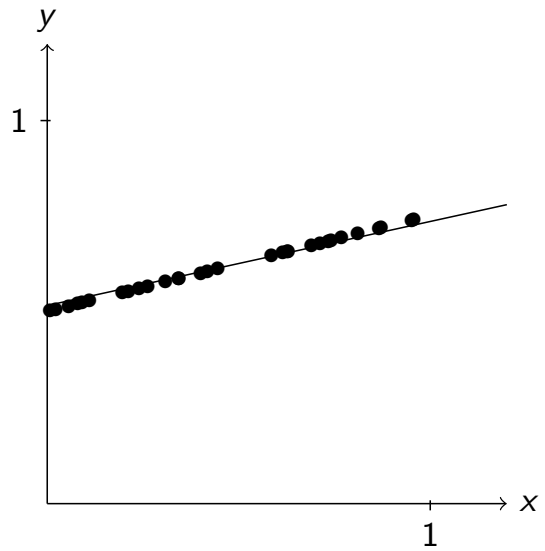
$$a = 0.21 \quad b = 0.52 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000131$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

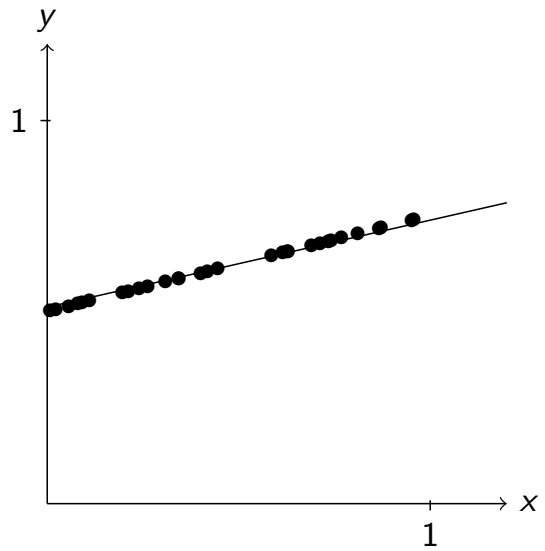
$$a = 0.22 \quad b = 0.52 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000073$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

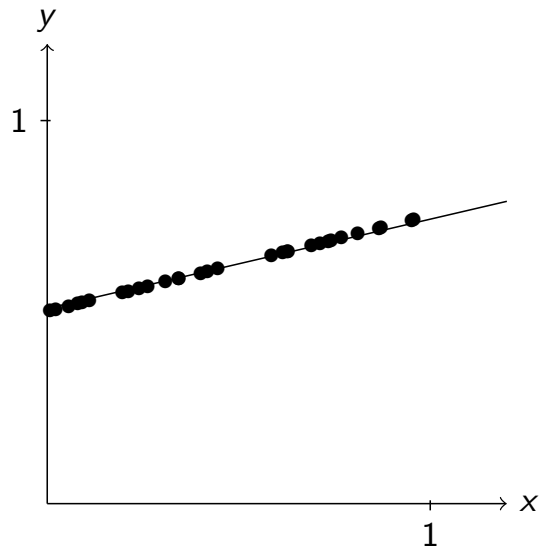
$$a = 0.23 \quad b = 0.51 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000040$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

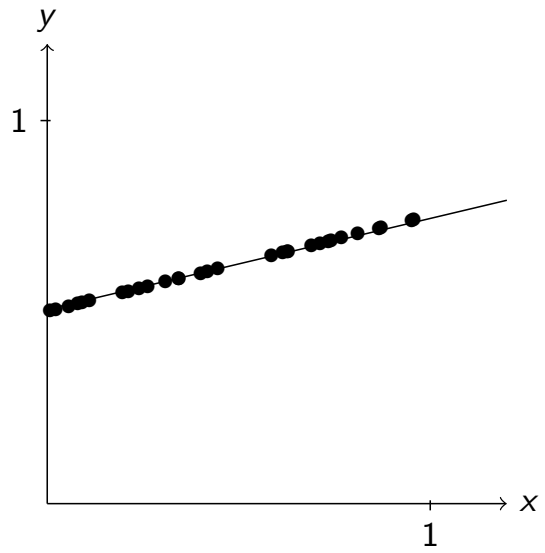
$$a = 0.23 \quad b = 0.51 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000022$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

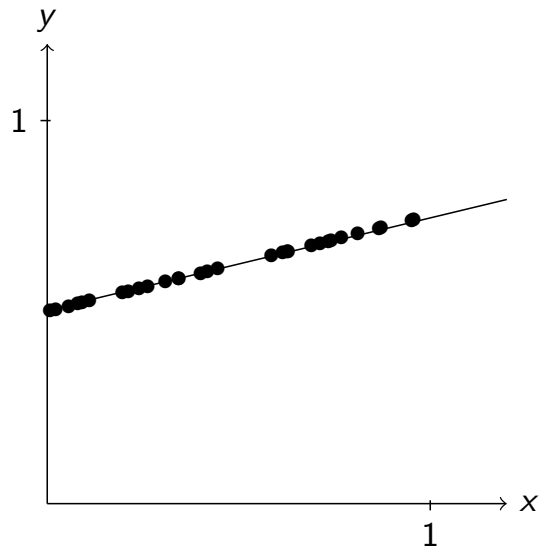
$$a = 0.24 \quad b = 0.51 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000012$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

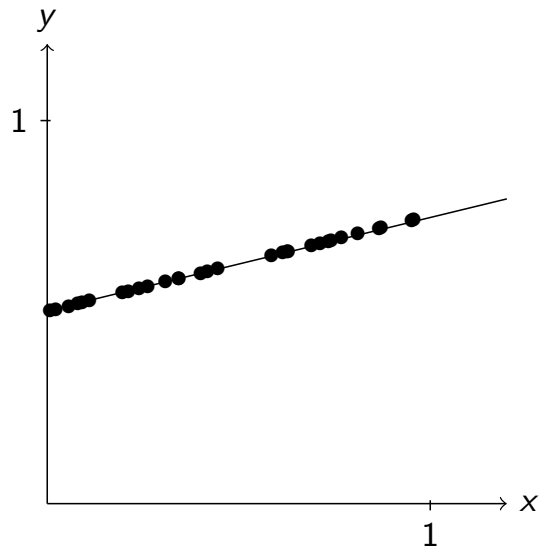
$$a = 0.24 \quad b = 0.51 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000007$$

And update a and b by substracting a small proportion of their gradient.

Parameter tuning

Gradient descent

We start with random values: $a = -2$ and $b = 1$



Then, we compute the average of the gradient along all (x, y) couples

$$a = 0.24 \quad b = 0.50 \quad \overline{\frac{dl}{da}} = 0.00 \quad \overline{\frac{dl}{db}} = 0.00 \quad \overline{l(x, y, a, b)} = 0.000004$$

And update a and b by substracting a small proportion of their gradient.