Lecture 1 : linear regression

By : Khalil idrissi

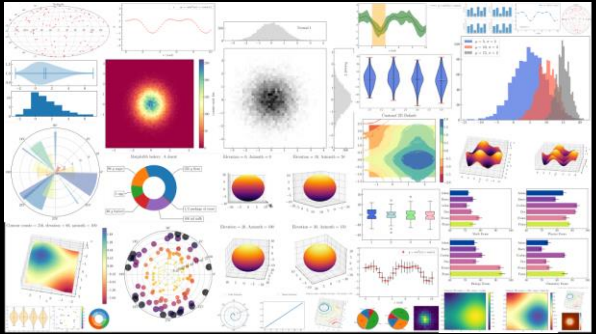
lecture 0 : intro
lecture 1 : linear regression
lecture 2 : SVMs
lecture 3 : dealing with images
lecture 4 : Neural network and backprop
lecture 5 : CNN, transfer learning and behavioral clonning
lecture 6 : autoencoders and image segmentation
lecture 7 : object detecion
lecture 8 : RNN ,LSTM, GRU
lecture 9 : decision trees, random forests, bagging, boosting, stacking
lecture 10 : Variational AE and GANs
lecture 11 : representation learning
lecture 12 : PCA and K-means clustering
lecture 13 : intro to Reinforcement learning

By : Khalil idrissi

lecture 0 : intro
lecture 1 : linear regression
lecture 2 : SVMs
lecture 3 : dealing with images
lecture 4 : Neural network and backprop
lecture 5 : CNN, transfer learning and behavioral clonning
lecture 6 : autoencoders and image segmentation
lecture 7 : object detecion
lecture 8 : RNN ,LSTM, GRU
lecture 9 : decision trees, random forests, bagging, boosting, stacking
lecture 10 : Variational AE and GANs
lecture 11 : representation learning
lecture 12 : PCA and K-means clustering
lecture 13 : intro to Reinforcement learning

By : Khalil idrissi

Today:



1- linear regression using OLS
2- linear regression using GD
3- working with sci-kit learn, numpy and matplotlib

By : Khalil idrissi

Recap

Recap

Machine learning → Approach for Ai

Recap

Machine learning → Approach for Ai

Dataset

Recap

Machine learning → Approach for Ai

Dataset

Recap

Machine learning → Approach for Ai

Dataset

- Training set
- Testing set

Recap

Machine learning → Approach for Ai

Dataset

- Training set
- Testing set

$$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N}$$

Recap

Machine learning → Approach for Ai

Dataset
- Training set
- Testing set

$$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N} \longrightarrow$$

Recap

Machine learning $\longrightarrow$ Approach for Ai

Dataset $\begin{cases} \text{Training set} \\ \text{Testing set} \end{cases}$

$$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N} \longrightarrow \text{Supervised learning}$$

Recap

Machine learning → Approach for Ai

Dataset
- Training set
- Testing set

$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N}$ → Supervised learning

$D = \{(x_i\ )\}_{\mathrm{i}=1}^{N}$

Recap

Machine learning $\longrightarrow$ Approach for Ai

Dataset $\begin{cases} \text{Training set} \\ \text{Testing set} \end{cases}$

$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N}$ $\longrightarrow$ Supervised learning

$D = \{(x_i\ )\}_{\mathrm{i}=1}^{N}$ $\longrightarrow$

Recap

Machine learning $\longrightarrow$ Approach for Ai

Dataset $\begin{cases} \text{Training set} \\ \text{Testing set} \end{cases}$

$$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N} \longrightarrow \text{Supervised learning}$$

$$D = \{(x_i\ )\}_{\mathrm{i}=1}^{N} \longrightarrow \text{Unsupervised learning}$$

Recap

Machine learning → Approach for Ai

Dataset
- Training set
- Testing set

$$D = \{(x_i, y_i)\}_{i=1}^{N} \longrightarrow \text{Supervised learning}$$

$$D = \{(x_i)\}_{i=1}^{N} \longrightarrow \text{Unsupervised learning}$$

if y_i is real valued variable , then we are talking about regression

Recap

Machine learning $\longrightarrow$ Approach for Ai

Dataset: Training set, Testing set

$$D = \{(x_i, y_i)\}_{i=1}^{N} \longrightarrow \text{Supervised learning}$$

$$D = \{(x_i)\}_{i=1}^{N} \longrightarrow \text{Unsupervised learning}$$

if y_i is real valued variable , then we are talking about regression

if y_i is belongs to a set of classes , then we are talking about classification

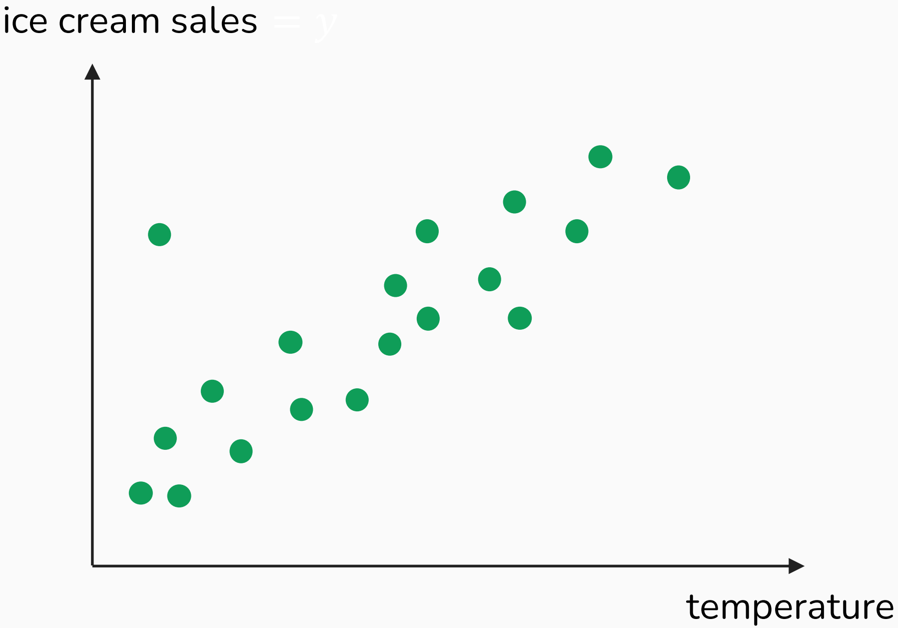
Remember ?

Remember ?

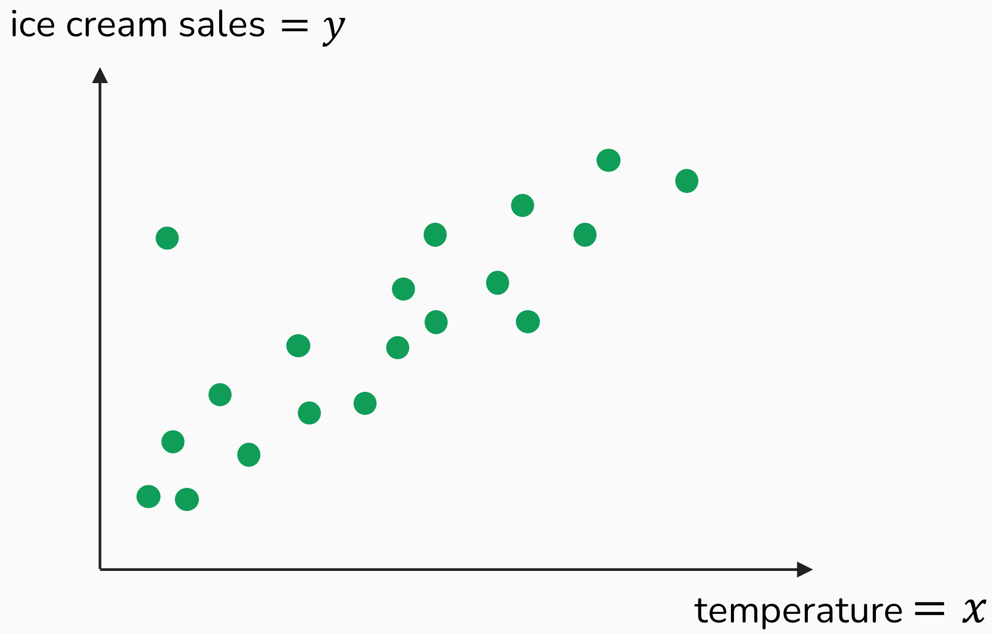
Remember ?

Linear regression

1- linear regression

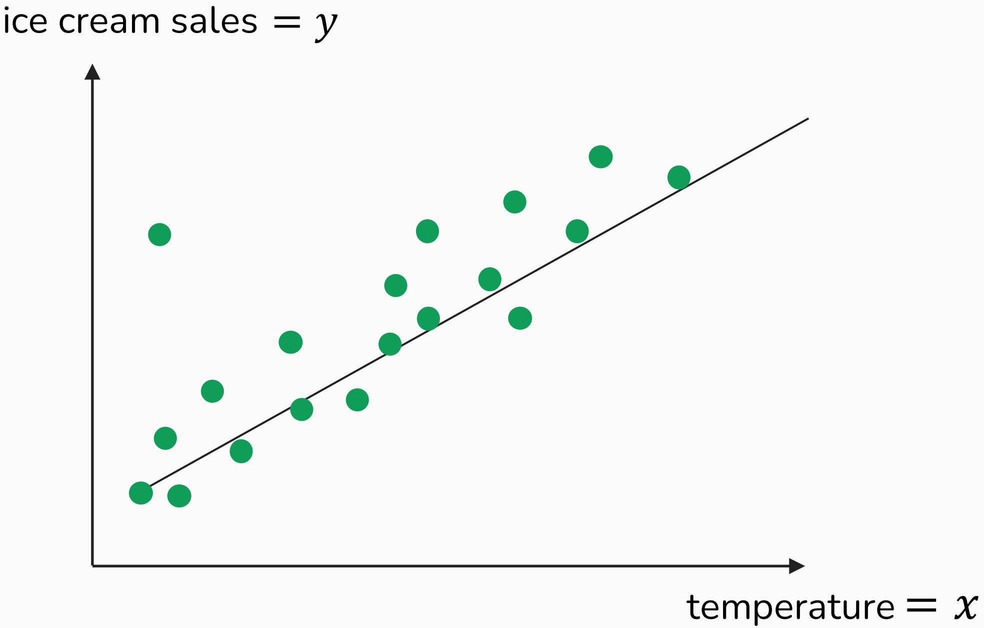


1- linear regression



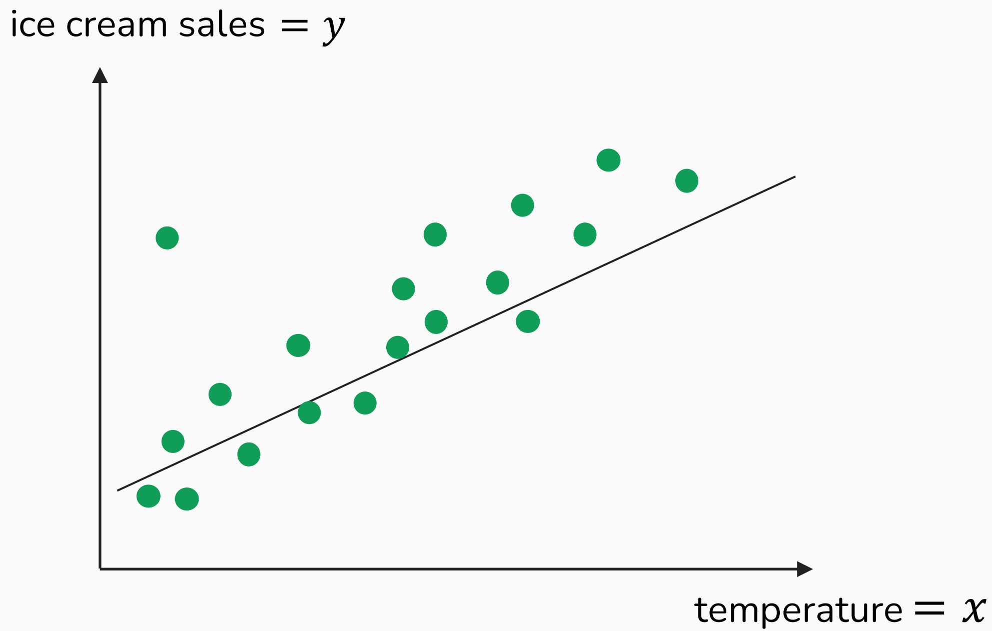
1- linear regression

1- linear regression

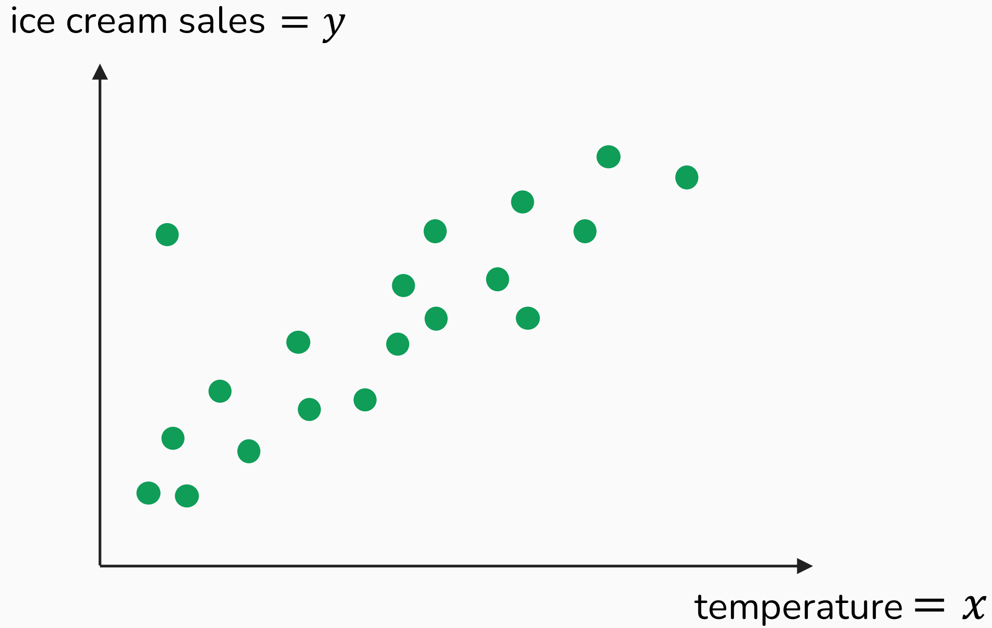


1- linear regression

1- linear regression

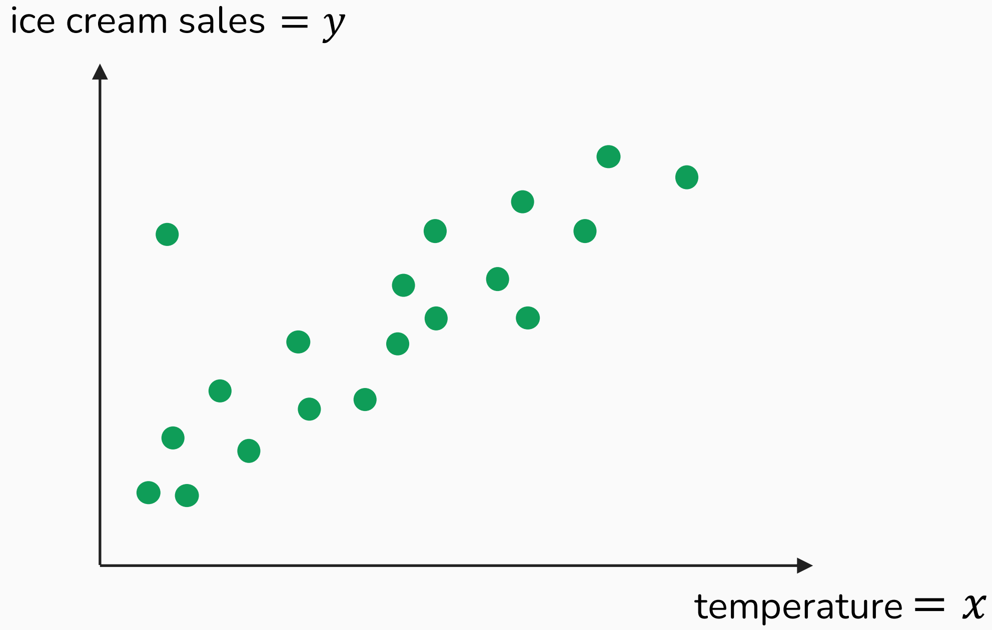


1- linear regression



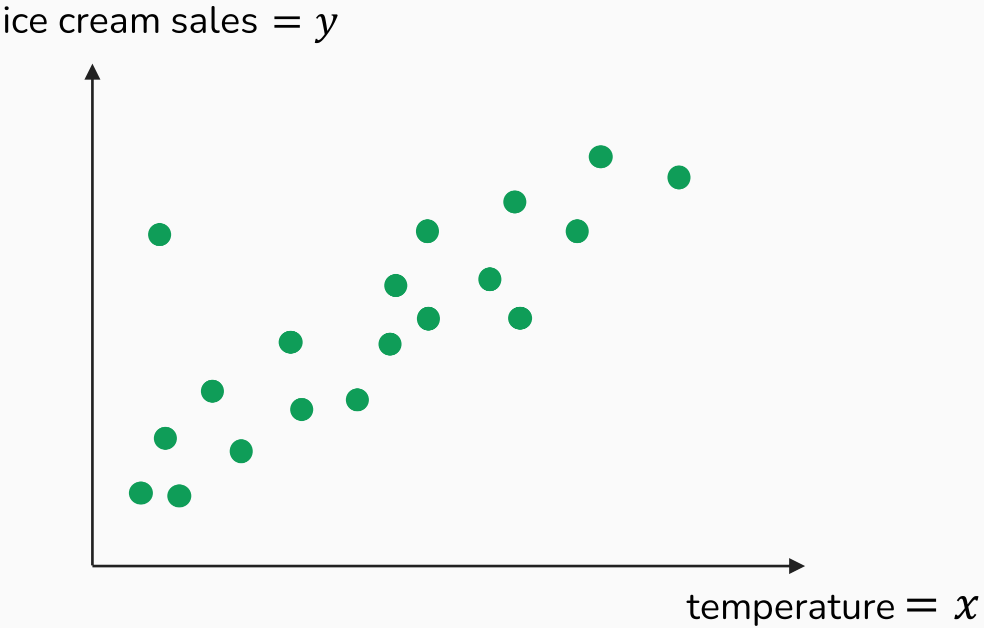
We need to find the best fit line

1- linear regression



1- using OLS (Ordinary Least Squares):

1- linear regression



1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

1- linear regression

1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

1- linear regression

1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

$$r_i{}^2 = (y_i - \hat{y}_i)^2$$

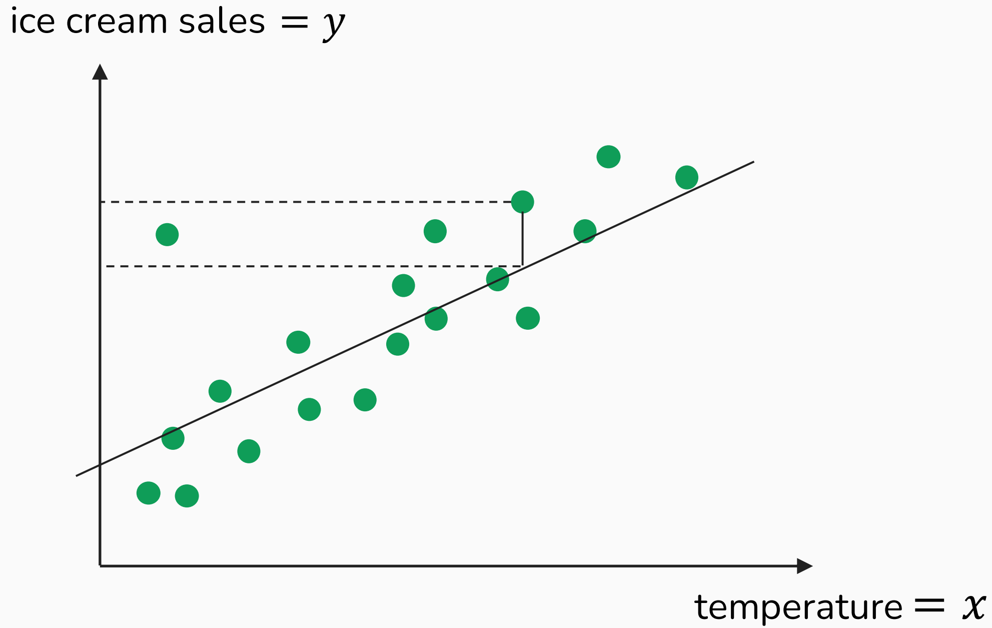
1- linear regression

1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

$$r_i^{\,2} = (y_i - \hat{y}_i)^2$$

1- linear regression

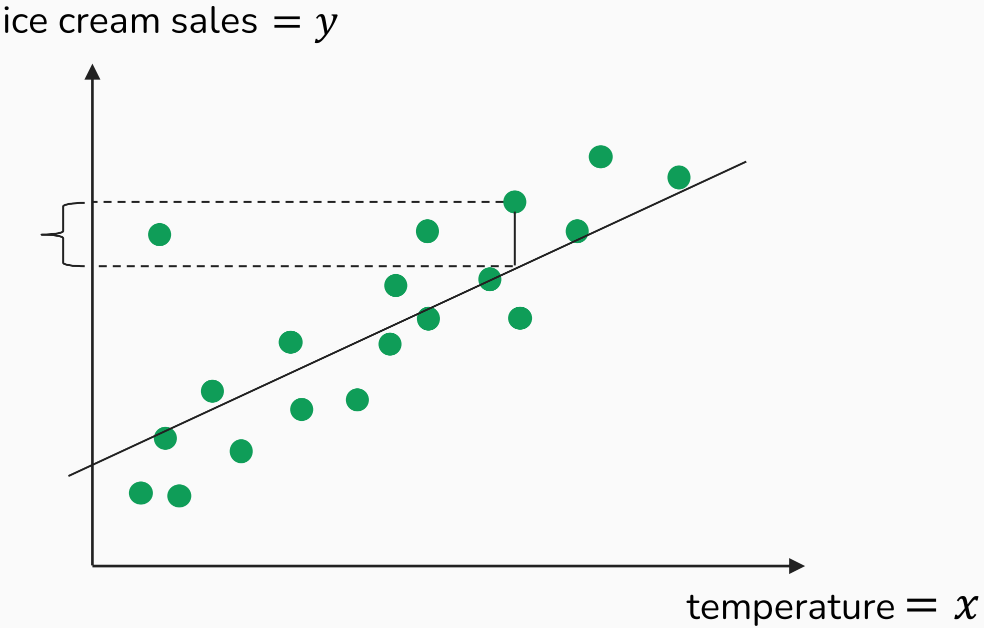


1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

$$r_i^{\,2} = (y_i - \hat{y}_i)^2$$

1- linear regression

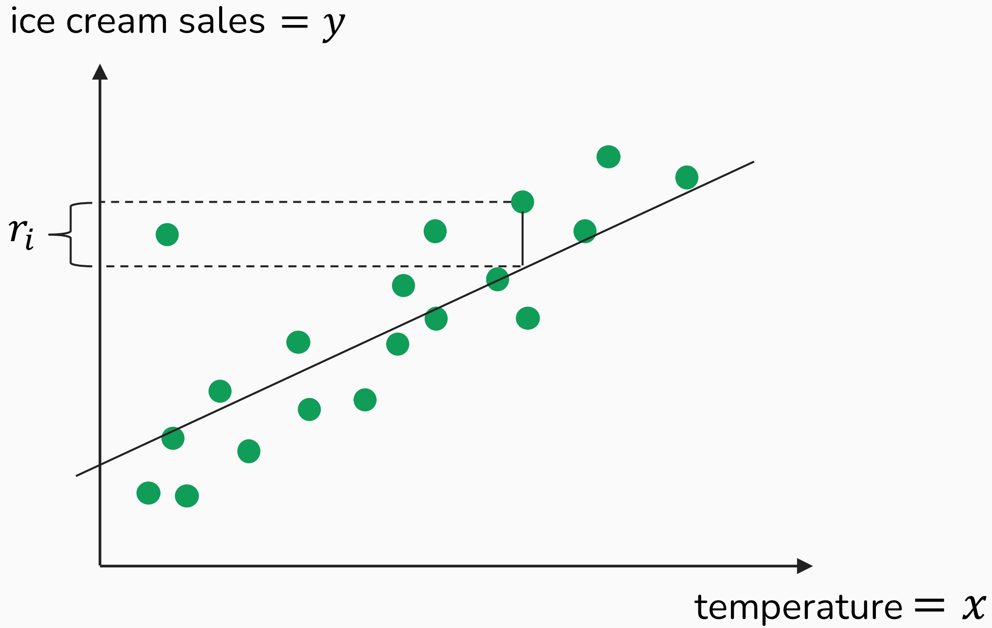


1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

$$r_i{}^2 = (y_i - \hat{y}_i)^2$$

1- linear regression



1- using OLS (Ordinary Least Squares):

Find a and b that minimizes the sum of squared residuals (SSR):

$$r_i^{\,2} = (y_i - \hat{y}_i)^2$$

We are trying to find a and b of the predicted line using OLS

We are trying to find a and b of the predicted line using OLS

The predicted line :

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

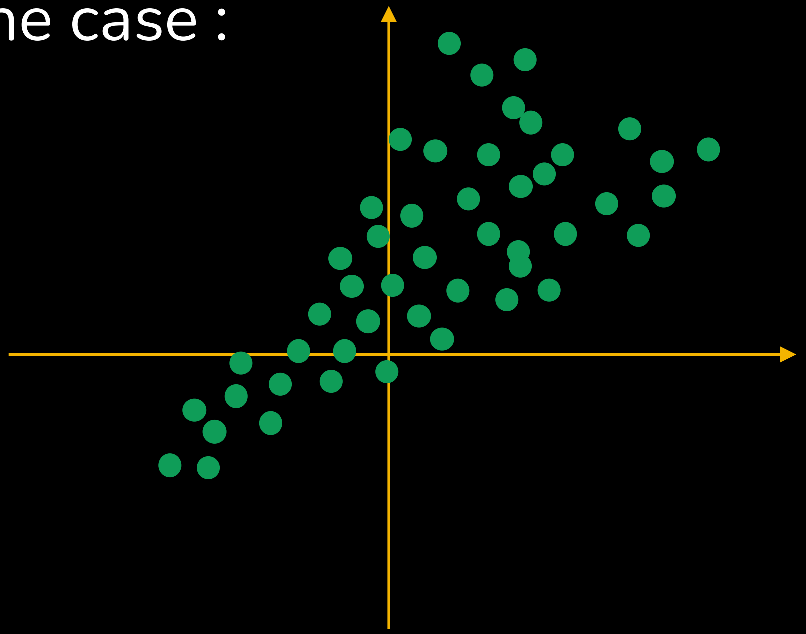
We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b + \varepsilon$

We often assume that ε has a gaussian distribution : $N(0, \sigma^2)$

→ A constant variance aka homoscedasticity

If it is not the case :



→ the linear regression model is likely to give incorrect estimates

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

Minimize :

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

Minimize : $r_i{}^2 = (y_i - \hat{y}_i)^2$ for all i

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

Minimize : $r_i{}^2 = (y_i - \hat{y}_i)^2$ for all i

$\longrightarrow$

We are trying to find a and b of the predicted line using OLS

The predicted line : $\hat{y}(x) = ax + b$

Minimize : $r_i{}^2 = (y_i - \hat{y}_i)^2$ for all i

$$\longrightarrow \quad \min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - ax_i - b)^2 =$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - ax_i - b)^2 =$$

$$\min \sum_{i=1}^{n} [(y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2] =$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - ax_i - b)^2 =$$

$$\min \sum_{i=1}^{n} [(y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2] =$$

$$\min\{ \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + \sum_{i=1}^{n} b^2 \}$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - a x_i - b)^2 =$$

$$\min \sum_{i=1}^{n} [(y_i - a x_i)^2 - 2(y_i - a x_i) b + b^2] =$$

$$\min\{ \sum_{i=1}^{n} (y_i - a x_i)^2 - 2b \sum_{i=1}^{n} (y_i - a x_i) + \sum_{i=1}^{n} b^2 \}$$

$$\min\{ \sum_{i=1}^{n} (y_i - a x_i)^2 - 2b \sum_{i=1}^{n} (y_i - a x_i) + b^2 . n \}$$

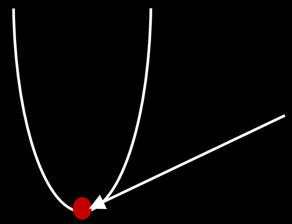
$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - ax_i - b)^2 =$$

$$\min \sum_{i=1}^{n} [(y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2] =$$

$$\min\{ \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + \sum_{i=1}^{n} b^2 \}$$

$$\min\{ \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + b^2 . n \}$$

To minimize $y = ax^2 + bx + c$:



$$\left(\frac{-b}{2a}, doesn't\ matter \right)$$

$$\min \sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n} (y_i - ax_i - b)^2 =$$

$$\min \sum_{i=1}^{n} [(y_i - ax_i)^2 - 2(y_i - ax_i)b + b^2] =$$

$$\min\{ \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + \sum_{i=1}^{n} b^2 \}$$

$$\min\{ \sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + b^2 . n \}$$

$\sum_{i=1}^{n} (y_i - ax_i)^2 - 2b \sum_{i=1}^{n} (y_i - ax_i) + b^2 . n$ is minimized when $b = \frac{-b'}{2a}$

À la fin on trouve w :

$$a = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^{n}(x_i - \bar{x})^2}$$

$$\begin{cases} \bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n} x_i \\ \bar{y} = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n} y_i \end{cases}$$

De même on trouve b :

$$b = \frac{1}{n}\sum_{i=0}^{n} y_i - \frac{1}{n}\sum_{i=0}^{n} a x_i$$

À la fin on trouve w :

$$a = \frac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^{n}(x_i - \bar{x})^2}$$

$$\begin{cases} \bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n} x_i \\ \bar{y} = \dfrac{1}{n}\displaystyle\sum_{i=0}^{n} y_i \end{cases}$$

De même on trouve b :

$$b = \frac{1}{n}\sum_{i=0}^{n} y_i - \frac{1}{n}\sum_{i=0}^{n} a x_i$$

In real life , there is more than one feature x
We could have many x

In real life , there is more than one feature x
We could have many x

$y(x) = w^T x + \varepsilon$ where w, ε, x, y are vectors

In real life , there is more than one feature x
We could have many x

$y(x) = w^T x + \varepsilon$ where w, ε, x, y are vectors

In real life , there is more than one feature x
We could have many x

$y(x) = w^T x + \varepsilon$ where w, ε, x, y are vectors

$$\longrightarrow \quad w^T = \left(x^T x\right)^{-1} x^T y$$

In real life , there is more than one feature x
We could have many x

$$y(x) = w^T x + \varepsilon$$ where w, ε, x, y are vectors

$$w^T = \left(x^T x\right)^{-1} x^T y$$

Closed form equation or normal equation

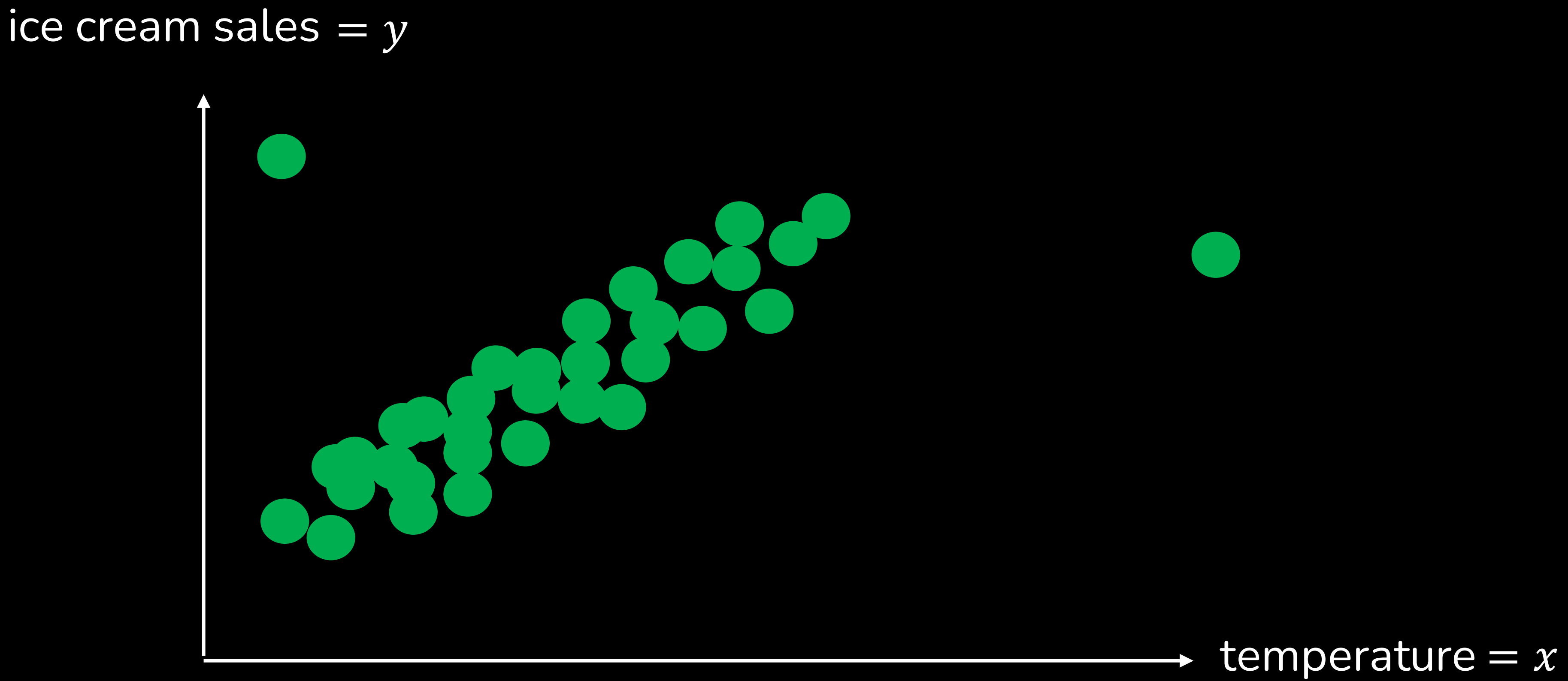
In real life , there is more than one feature x
We could have many x

$y(x) = w^T x + \varepsilon$ where w, ε, x, y are vectors

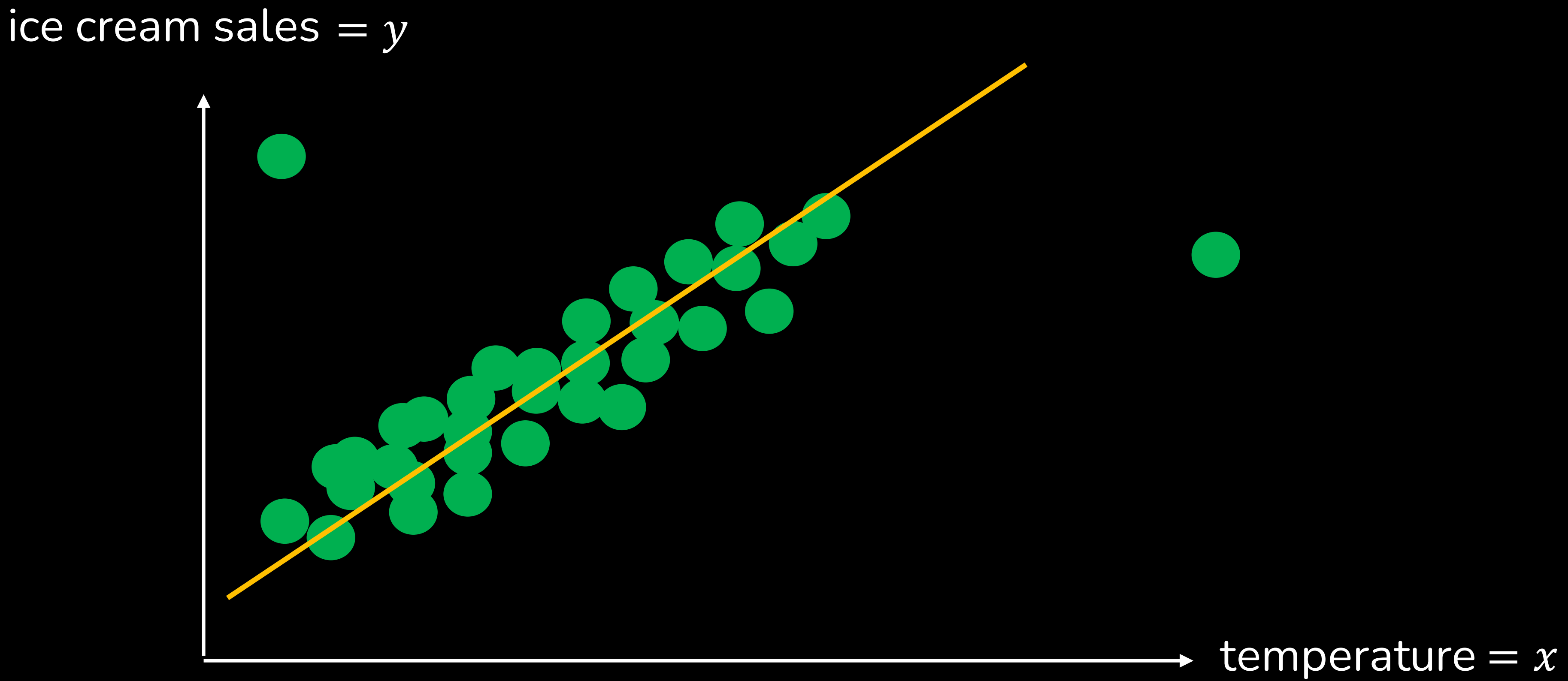
$$\boxed{w^T = \left(x^T x\right)^{-1} x^T y}$$

Closed form equation
or normal equation

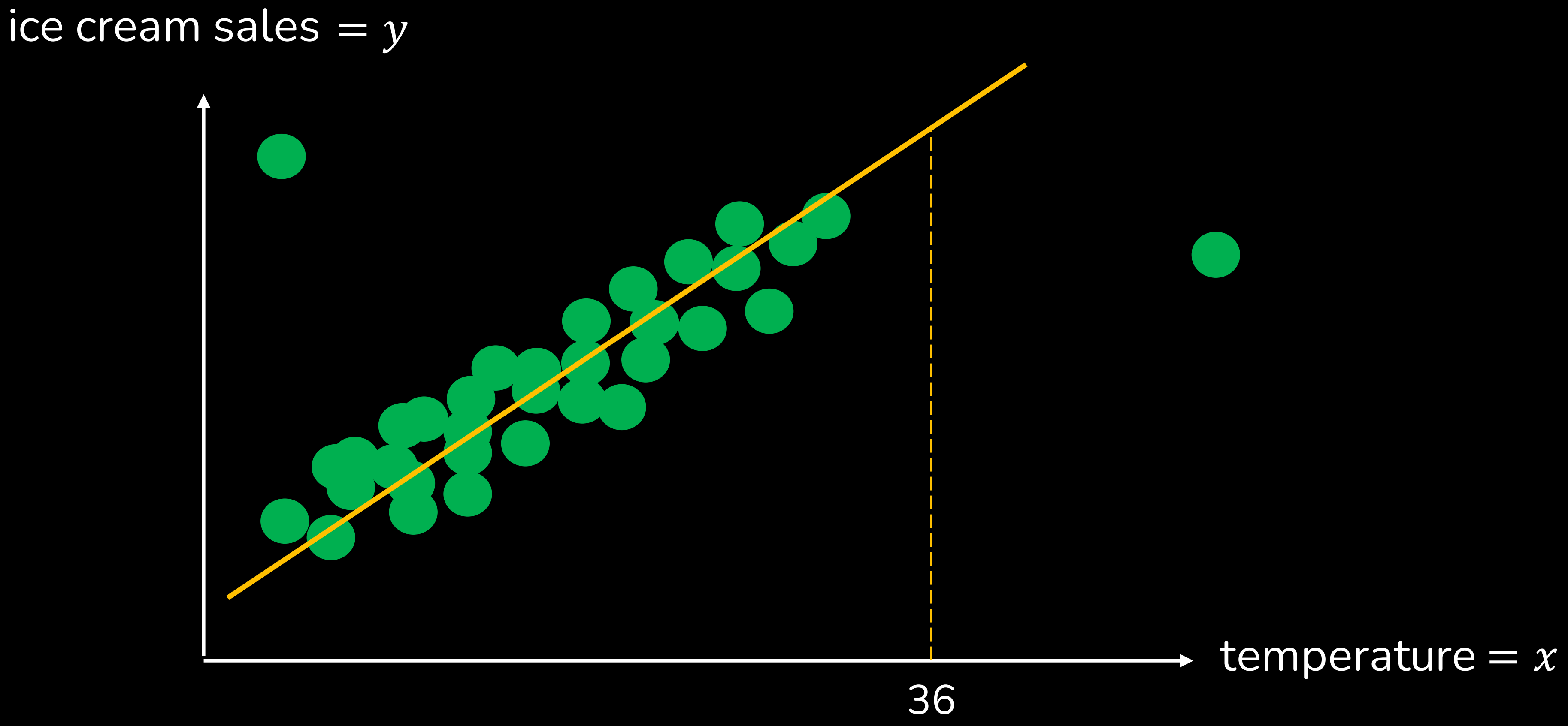
2- using gradient descent



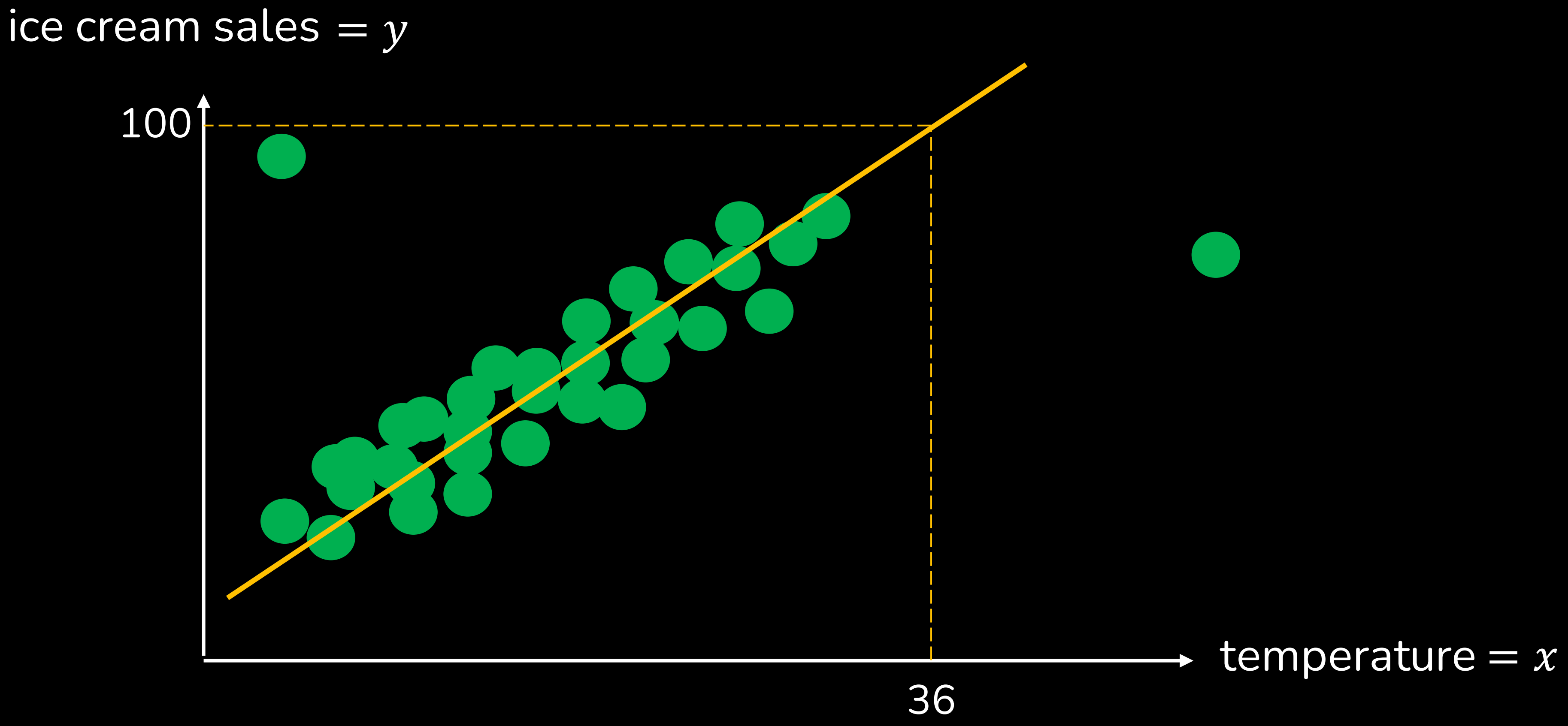
2- using gradient descent



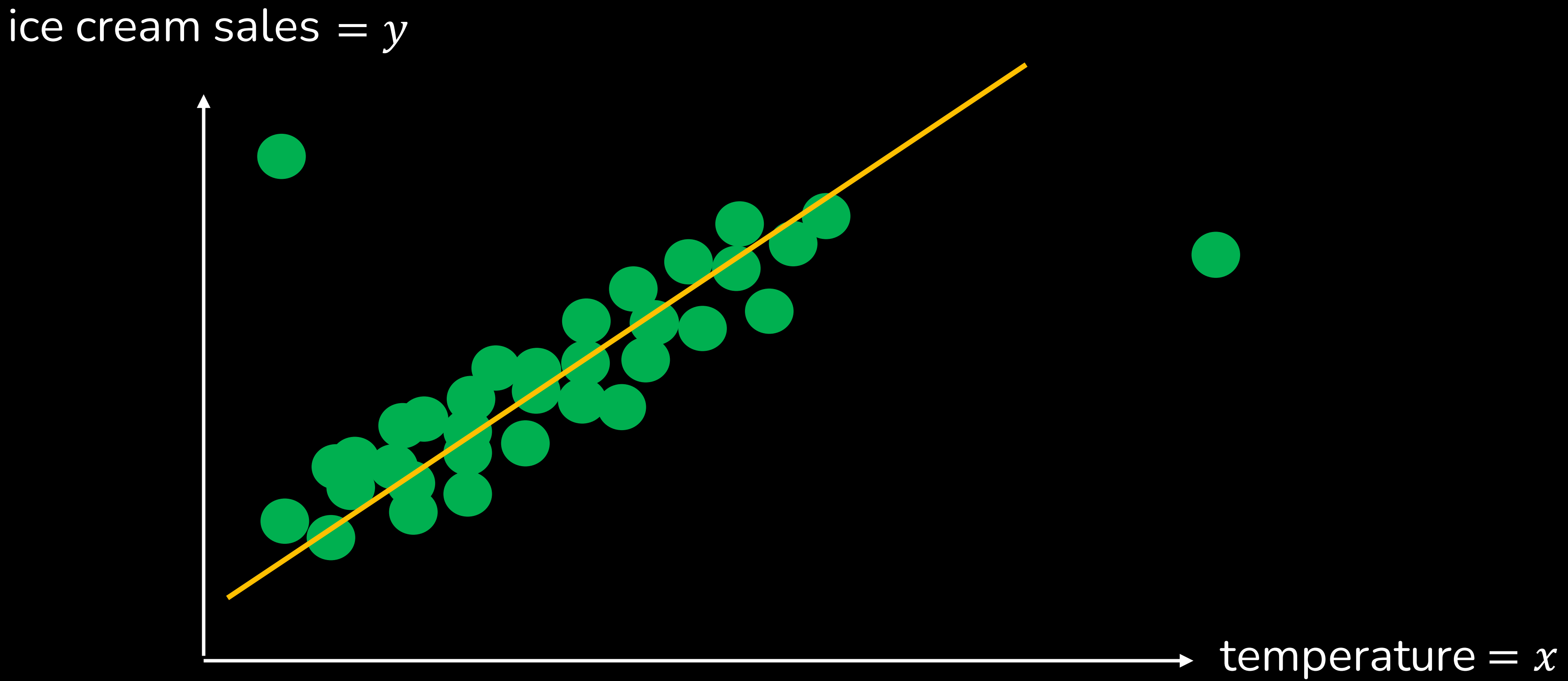
2- using gradient descent



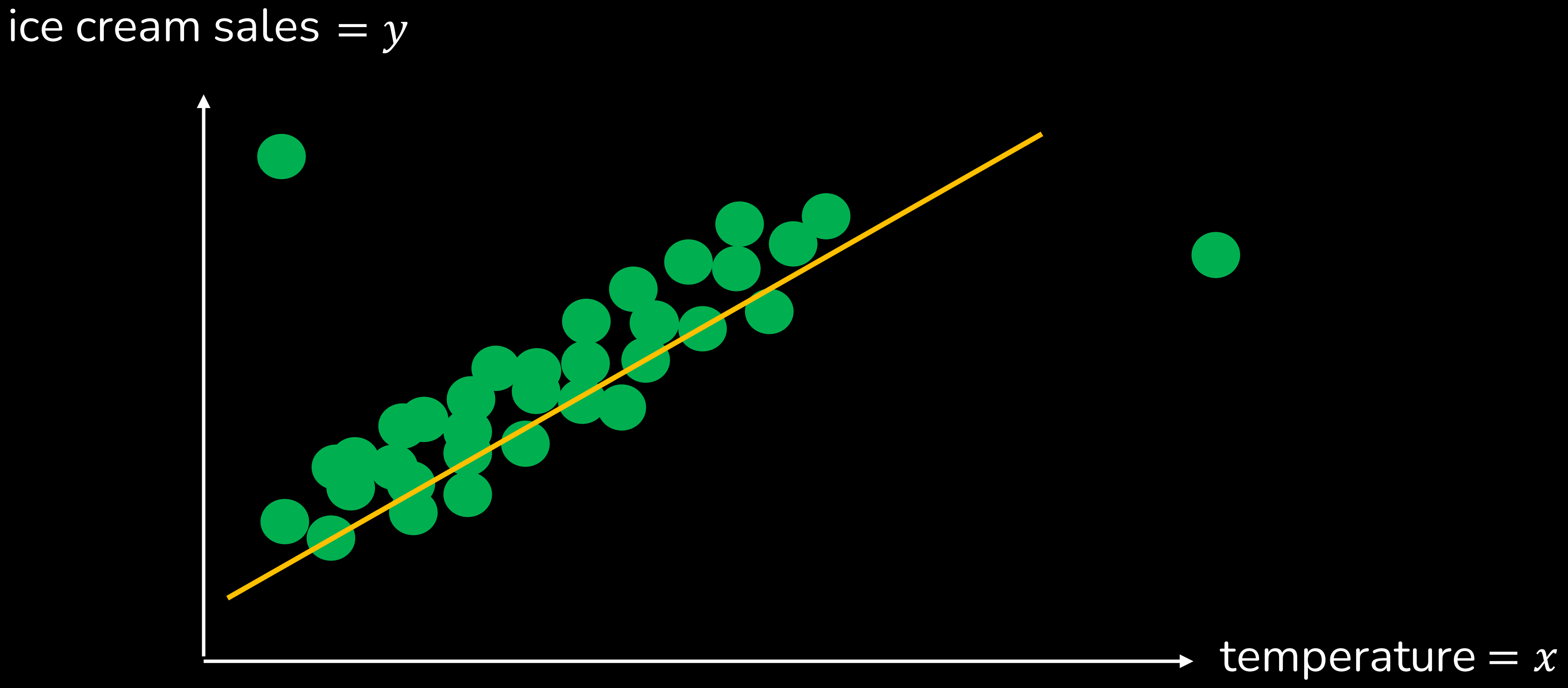
2- using gradient descent



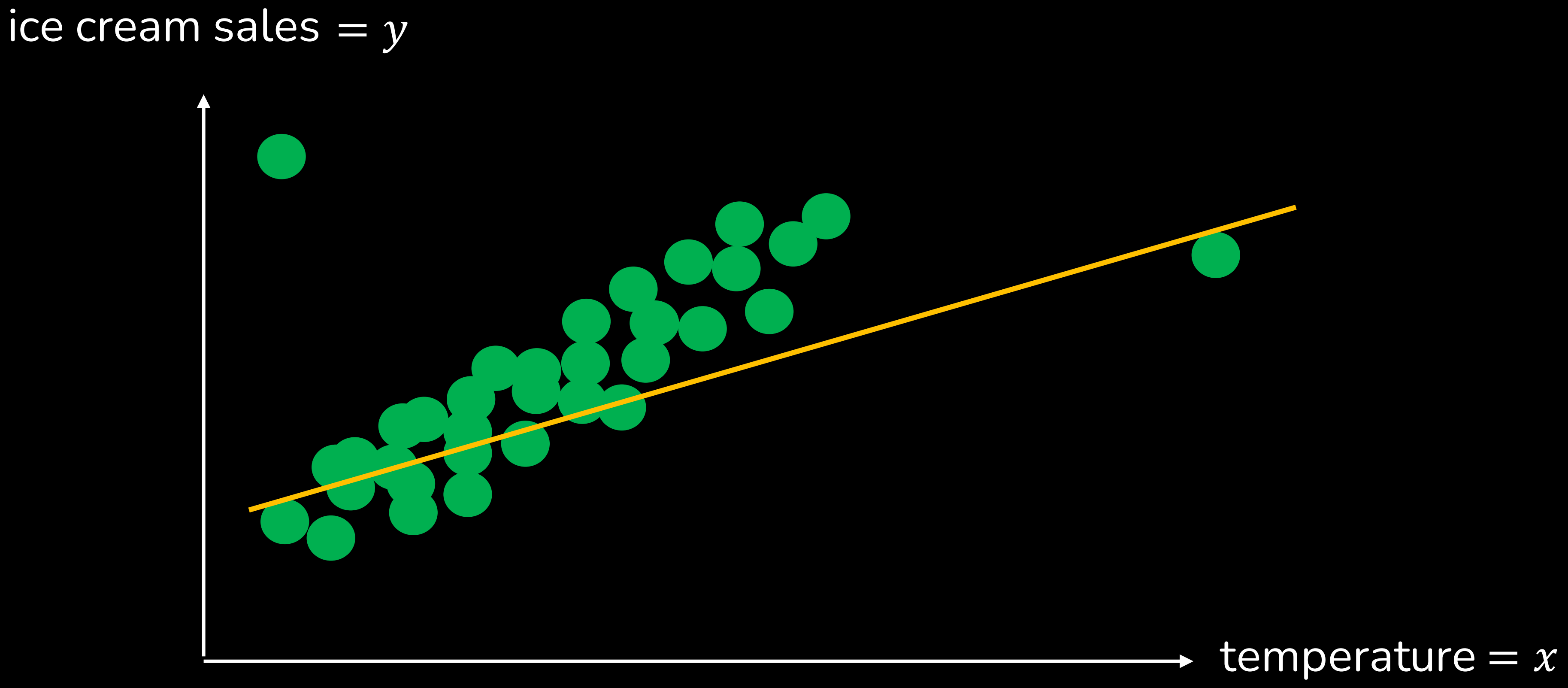
2- using gradient descent



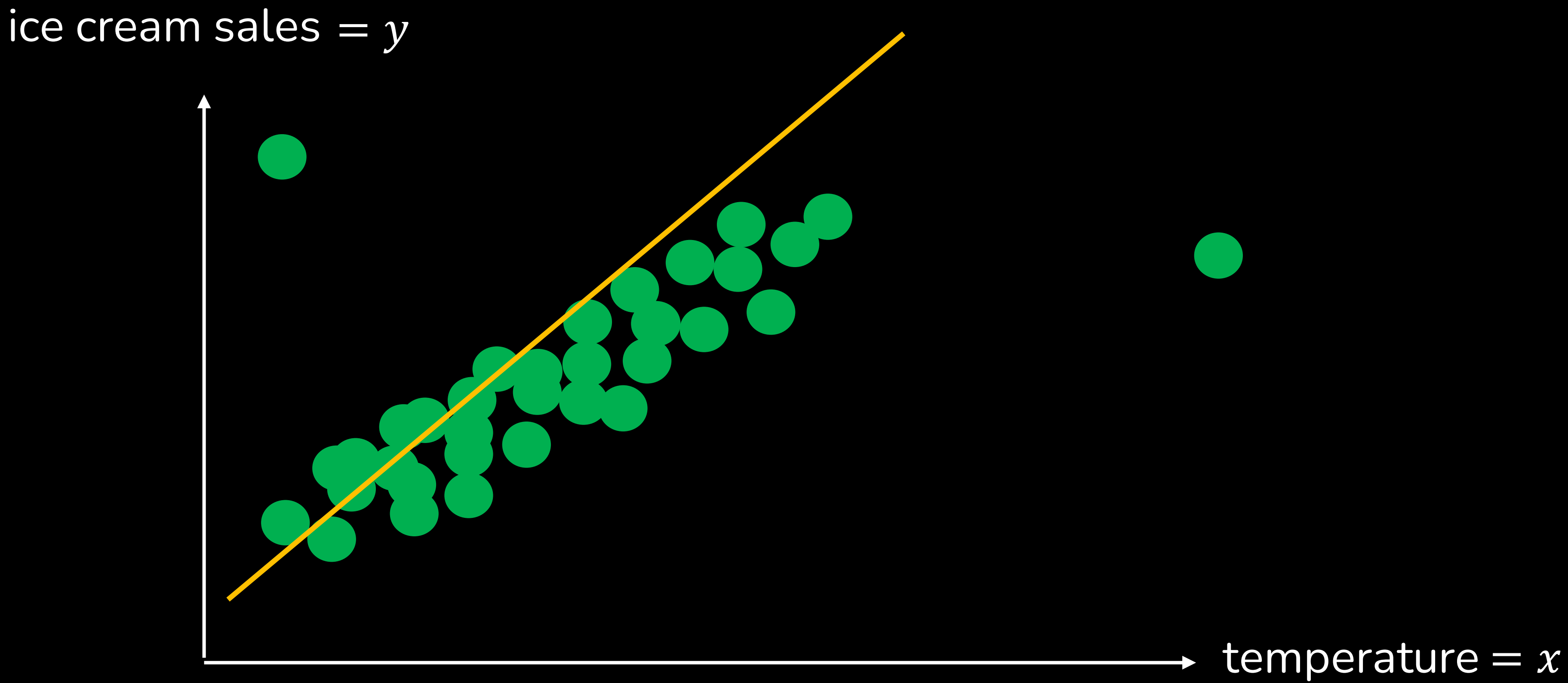
2- using gradient descent



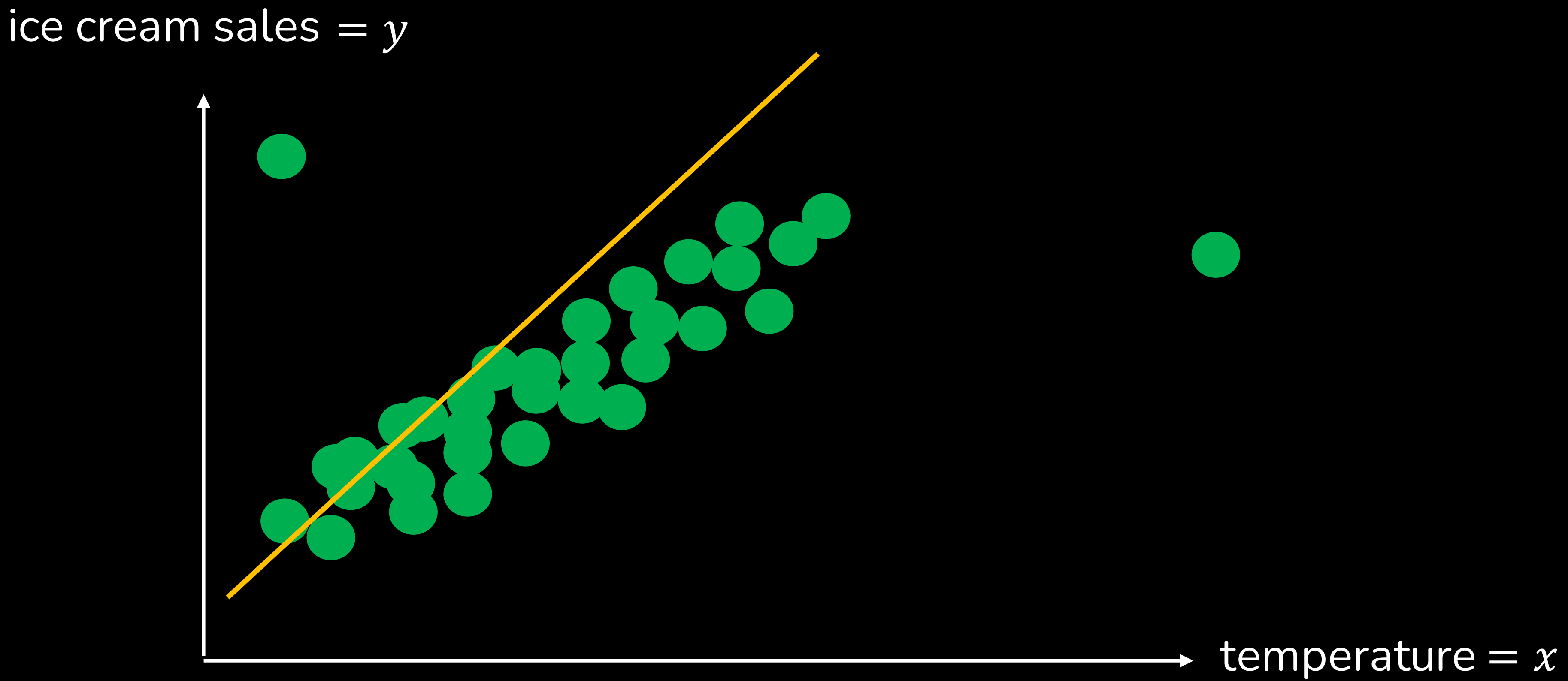
2- using gradient descent



2- using gradient descent



2- using gradient descent



Linear regression model

$$\hat{y}(x) = ax + b$$

$$D = \{(x_i, y_i)\}_{\mathrm{i}=1}^{N}$$

Linear regression model

$$\hat{y}(x) = ax + b$$

$$D = \{(x_i, y_i)\}_{\mathrm{i=1}}^{N} \longrightarrow \boxed{\begin{array}{c}\text{Linear regression}\\ \text{model}\\ \hat{y}(x) = ax + b\end{array}} \longrightarrow$$

$$D = \{(x_i, y_i)\}_{\mathrm{i=1}}^{N} \longrightarrow \boxed{\begin{array}{c}\text{Linear regression} \\ \text{model} \\ \hat{y}(x) = ax + b\end{array}} \longrightarrow \text{prediction}$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

→ Linear regression model

$$\hat{y}(x) = ax + b$$

→ prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

→

Linear regression model

$$\hat{y}(x) = ax + b$$

→ prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

→ **Linear regression model**

$$\hat{y}(x) = ax + b$$

→ prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2x + b$$

prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + b$$

prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

→ **Linear regression model**

$$\hat{y}(x) = 2.30 + 3$$

→ prediction

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = \frac{1}{n}\left[(y_1 - \hat{y}_1)^2 + \cdots\right]$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MSE = 180$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

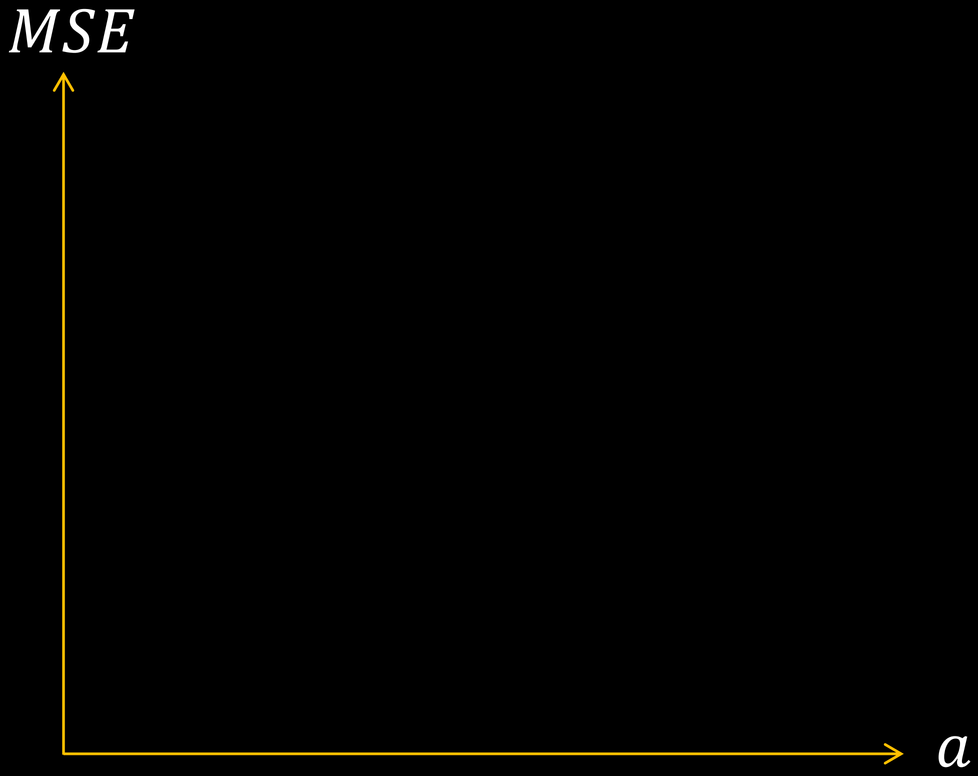
$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

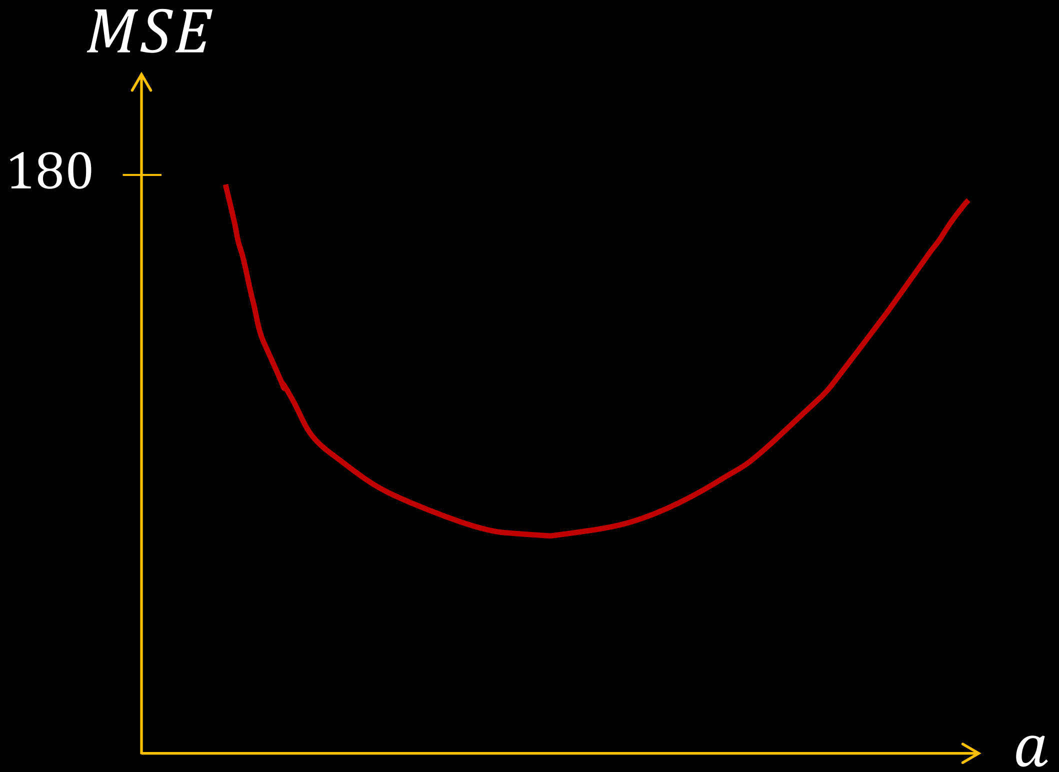
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



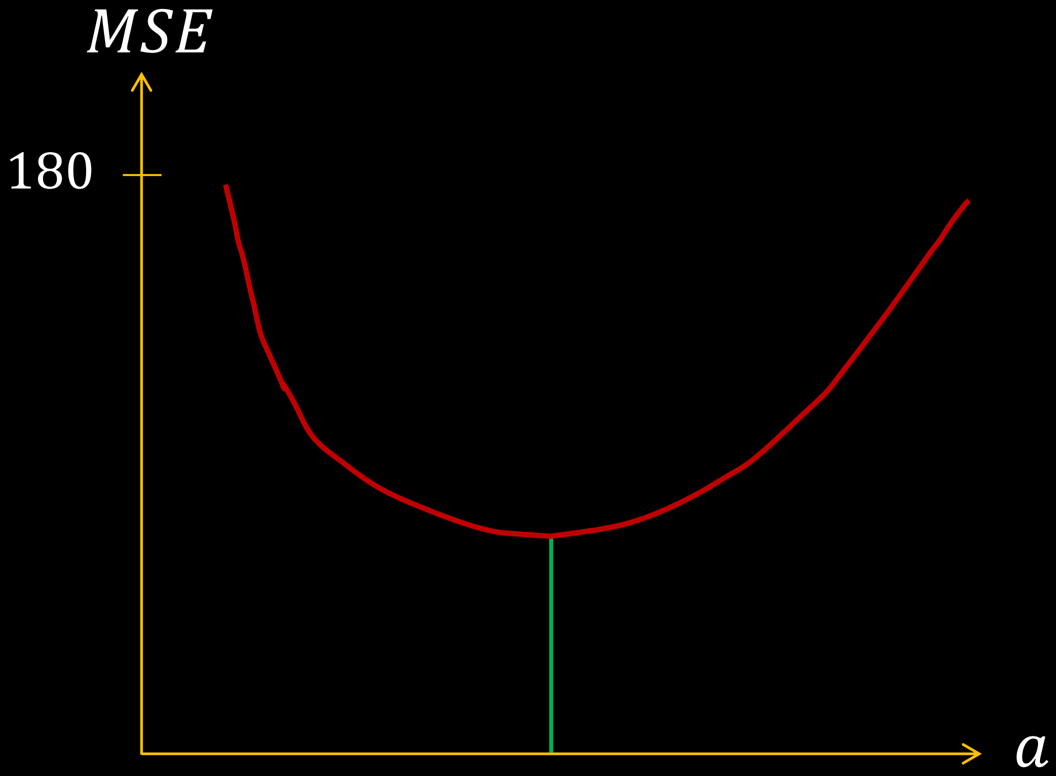
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



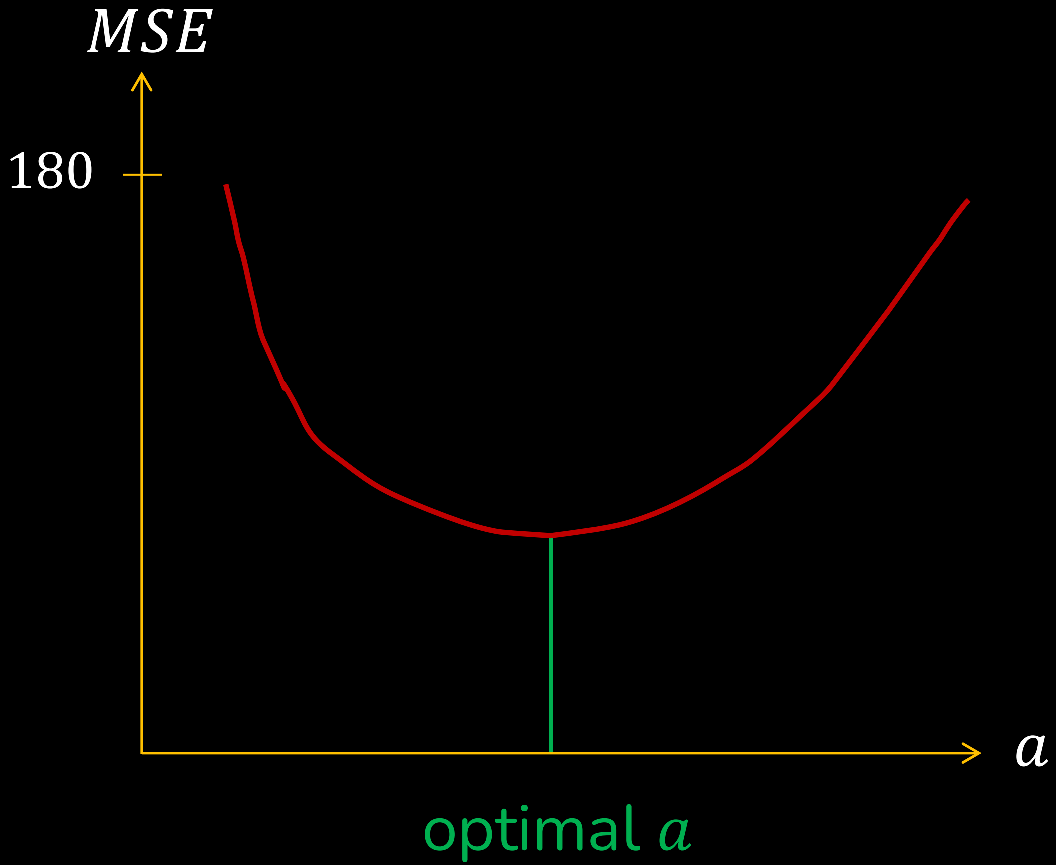
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



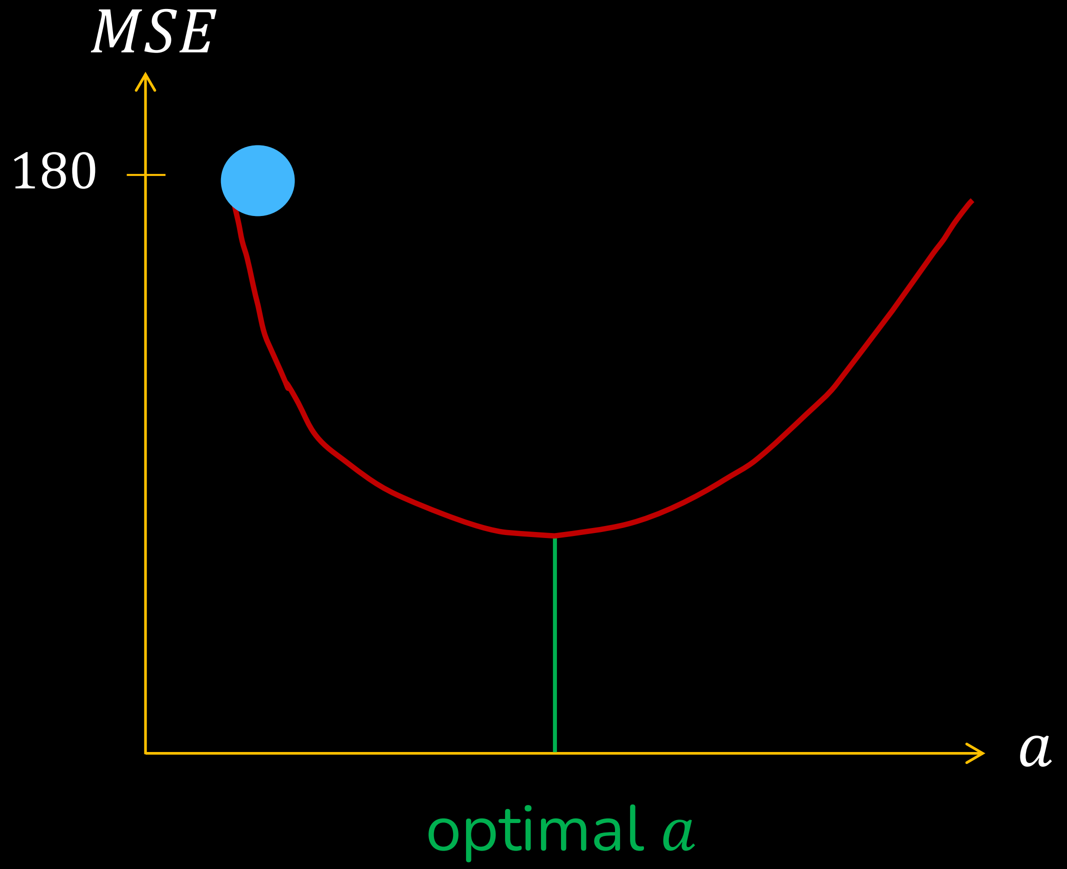
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



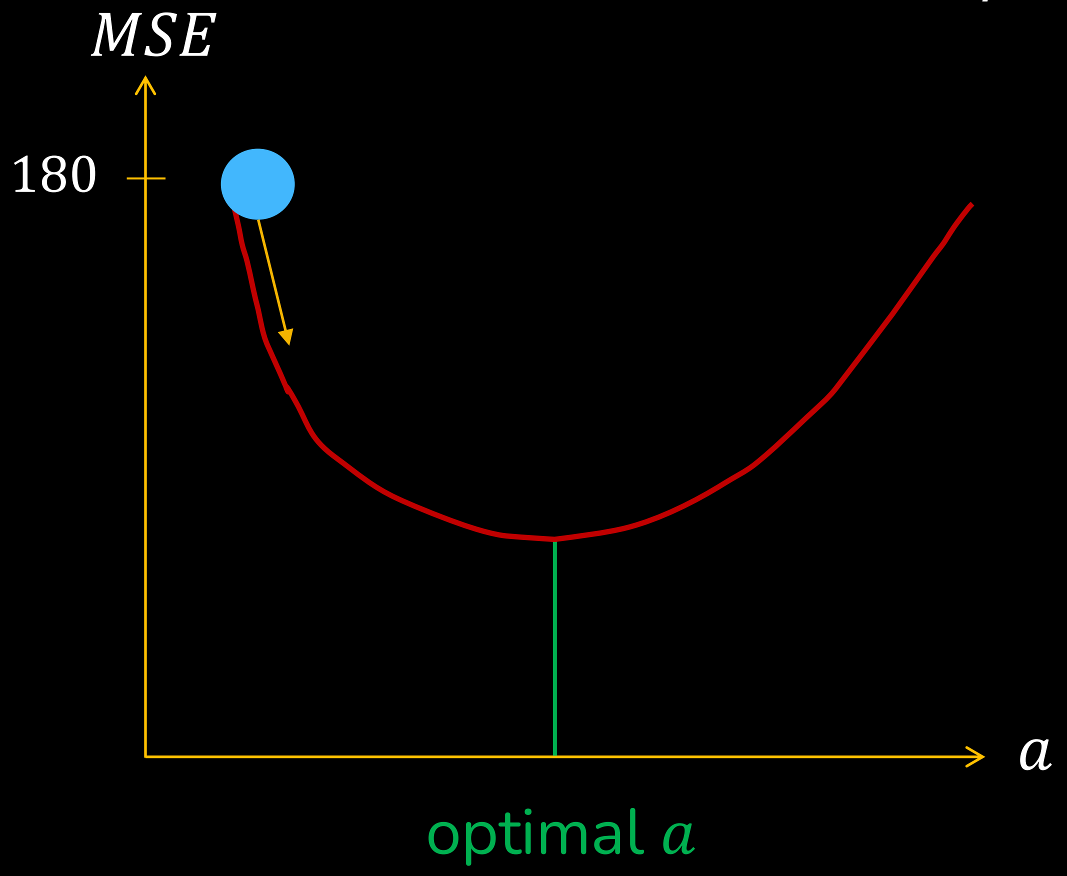
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$

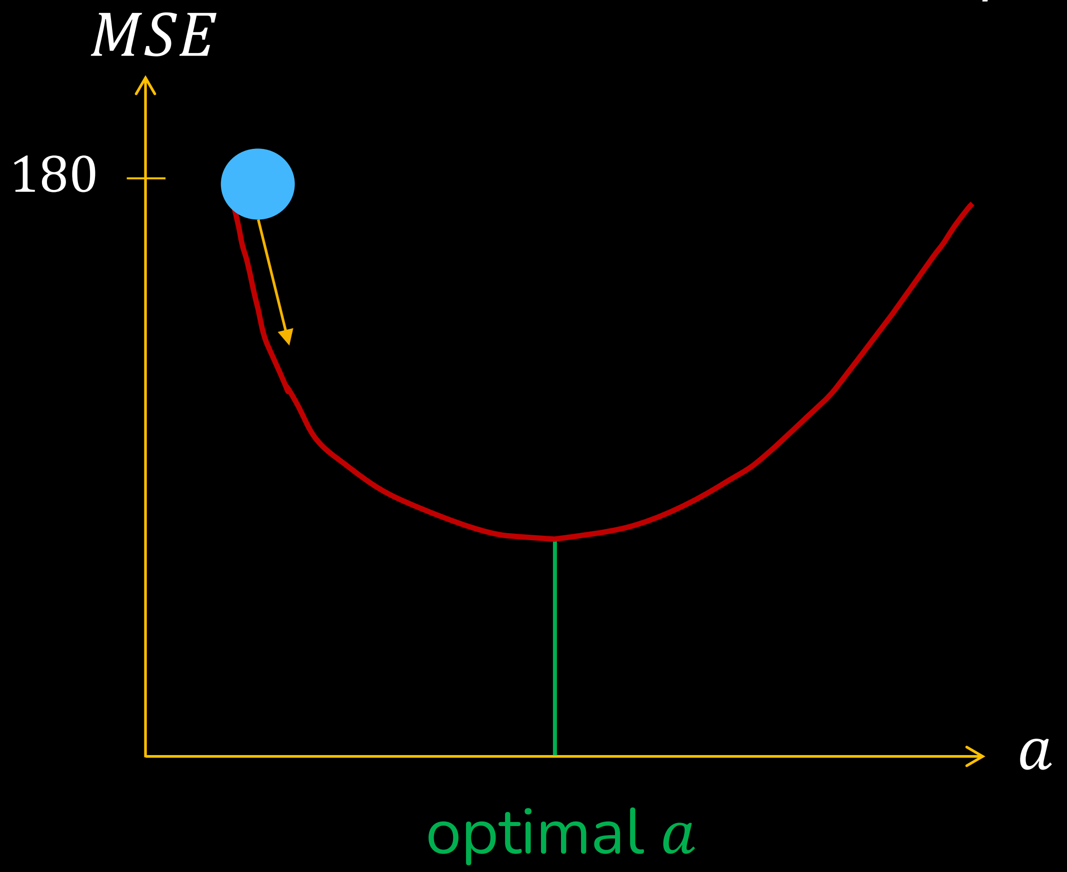


$$a = a - \eta \nabla MSE$$

$$MSE = 180$$

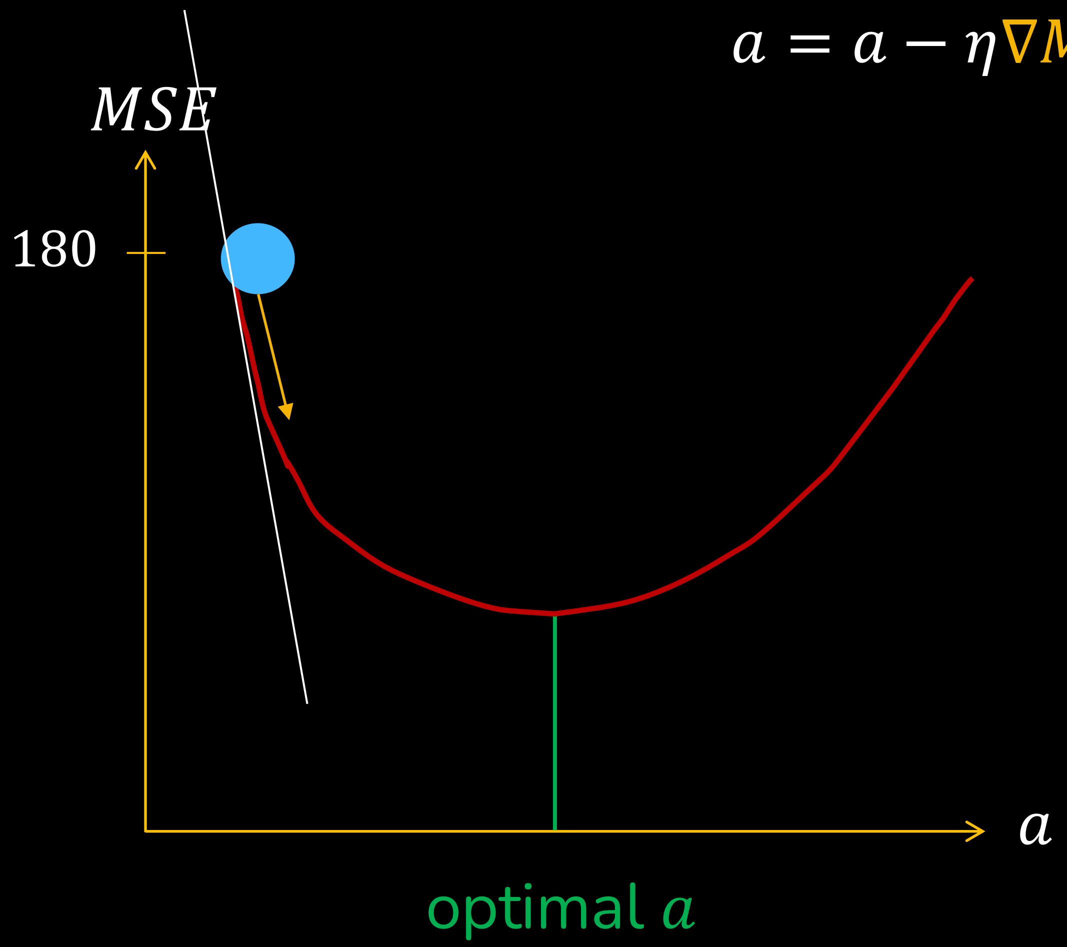


$$a = a - \eta \nabla MSE \qquad MSE = 180$$



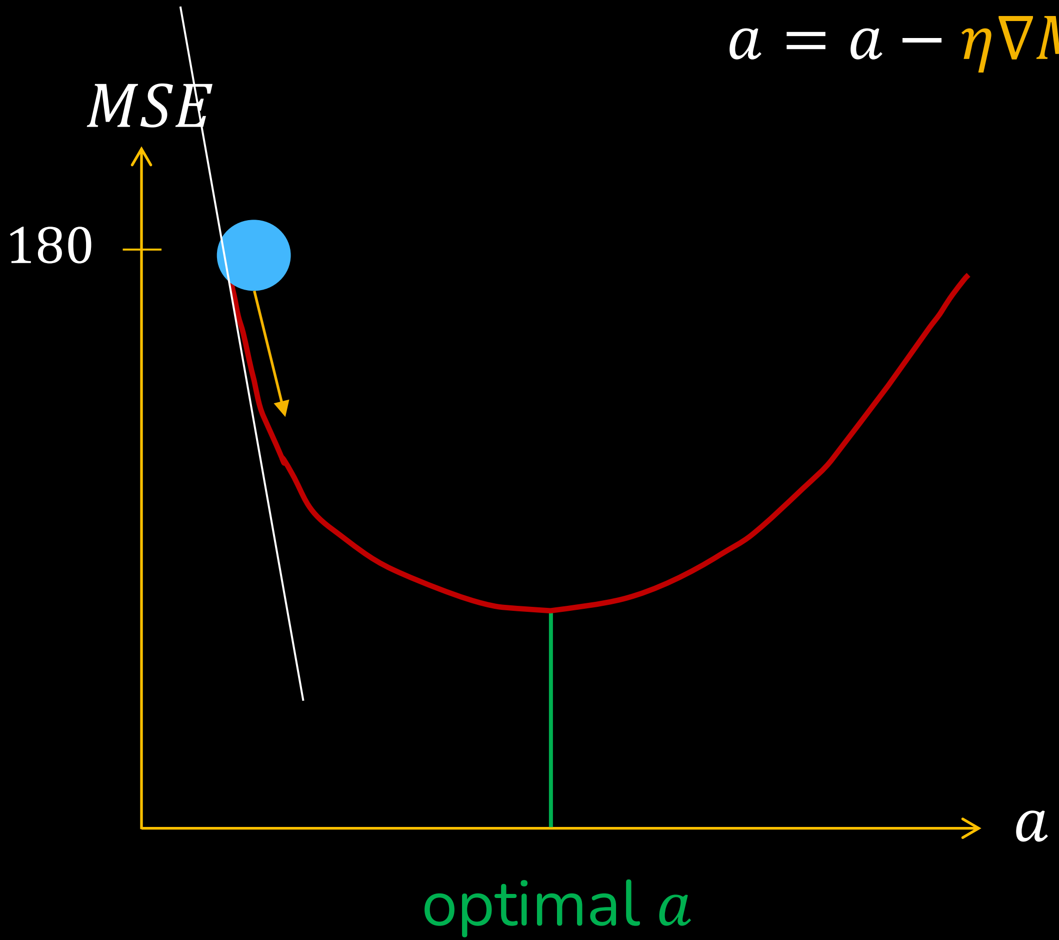
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



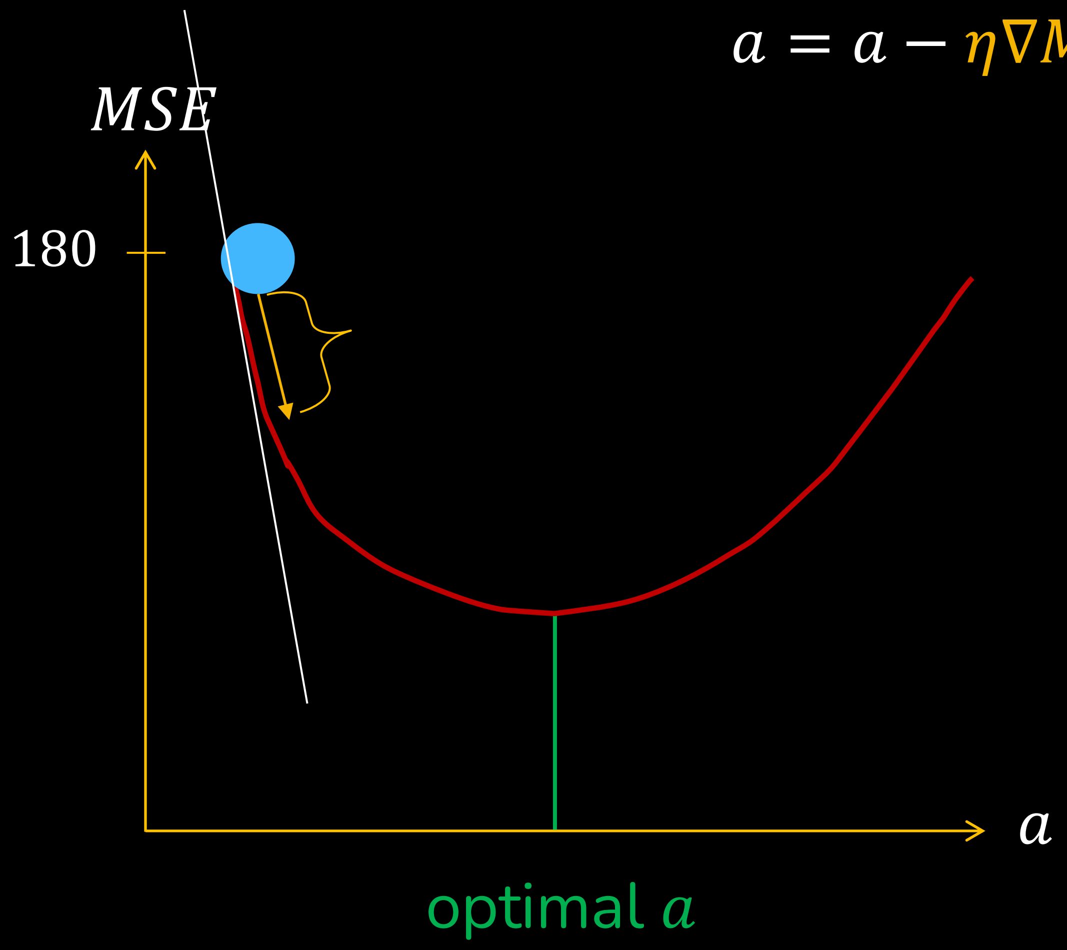
$$a = a - \eta \nabla MSE$$

$MSE = 180$



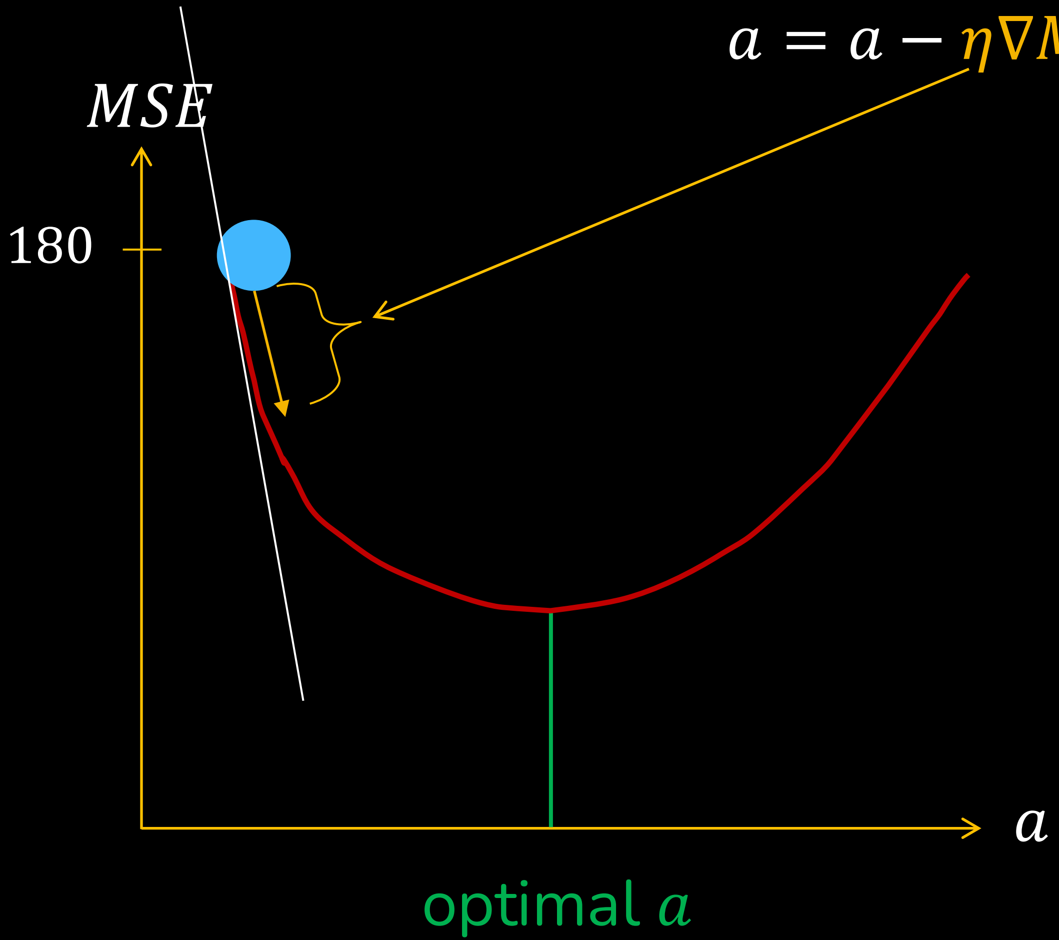
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



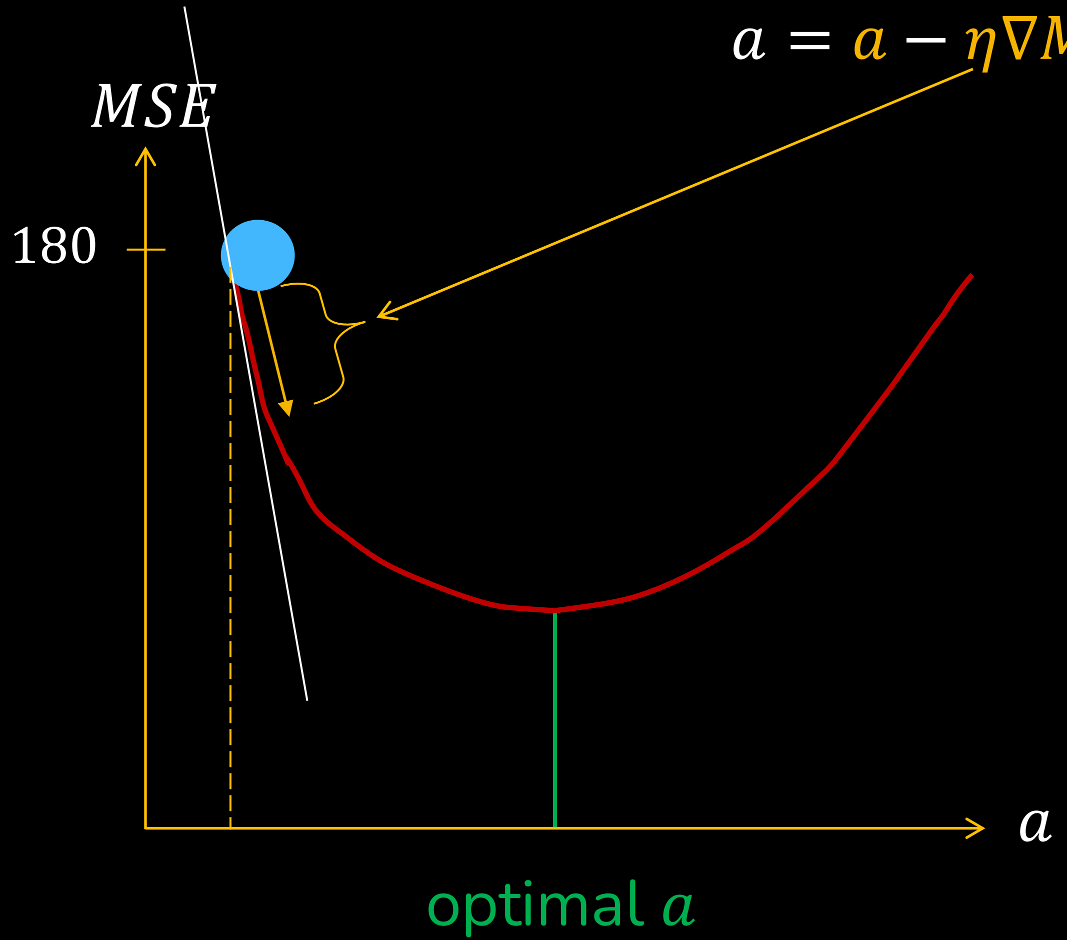
$$a = a - \eta\nabla MSE$$

$$MSE = 180$$



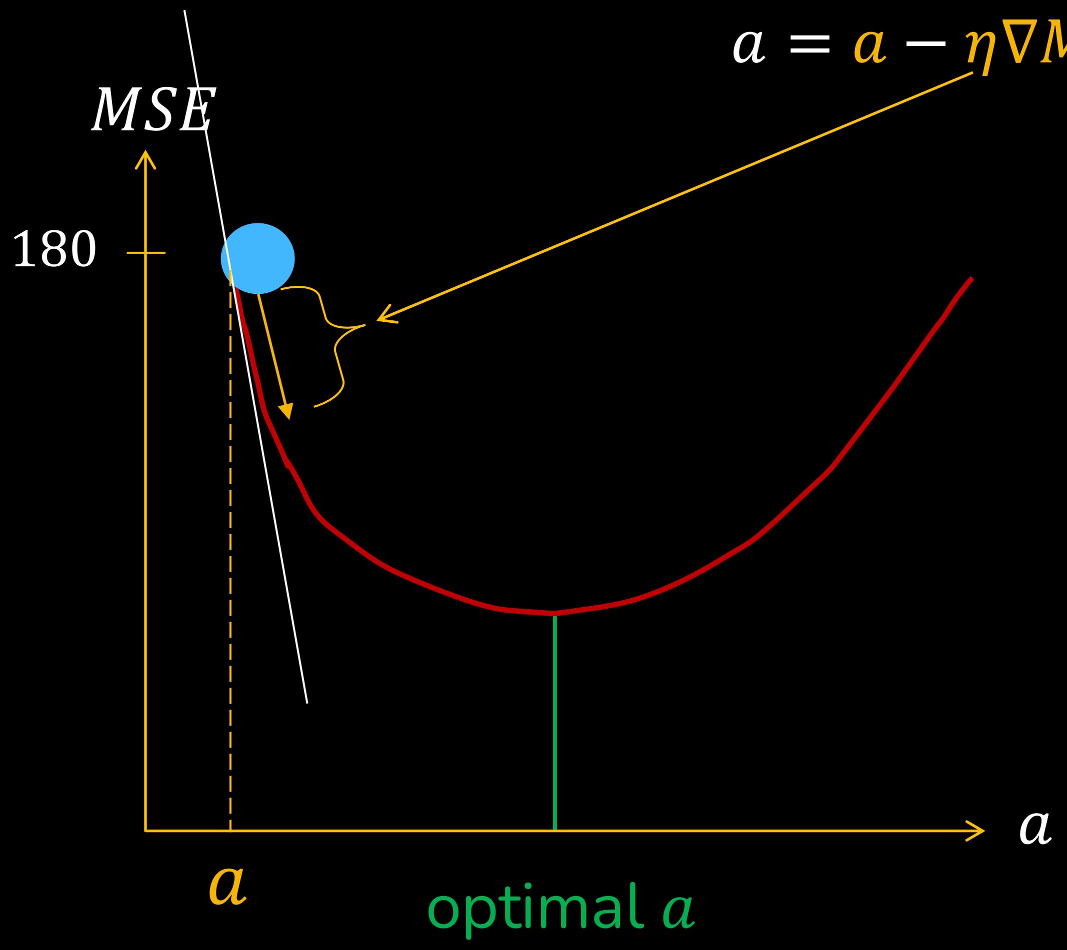
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



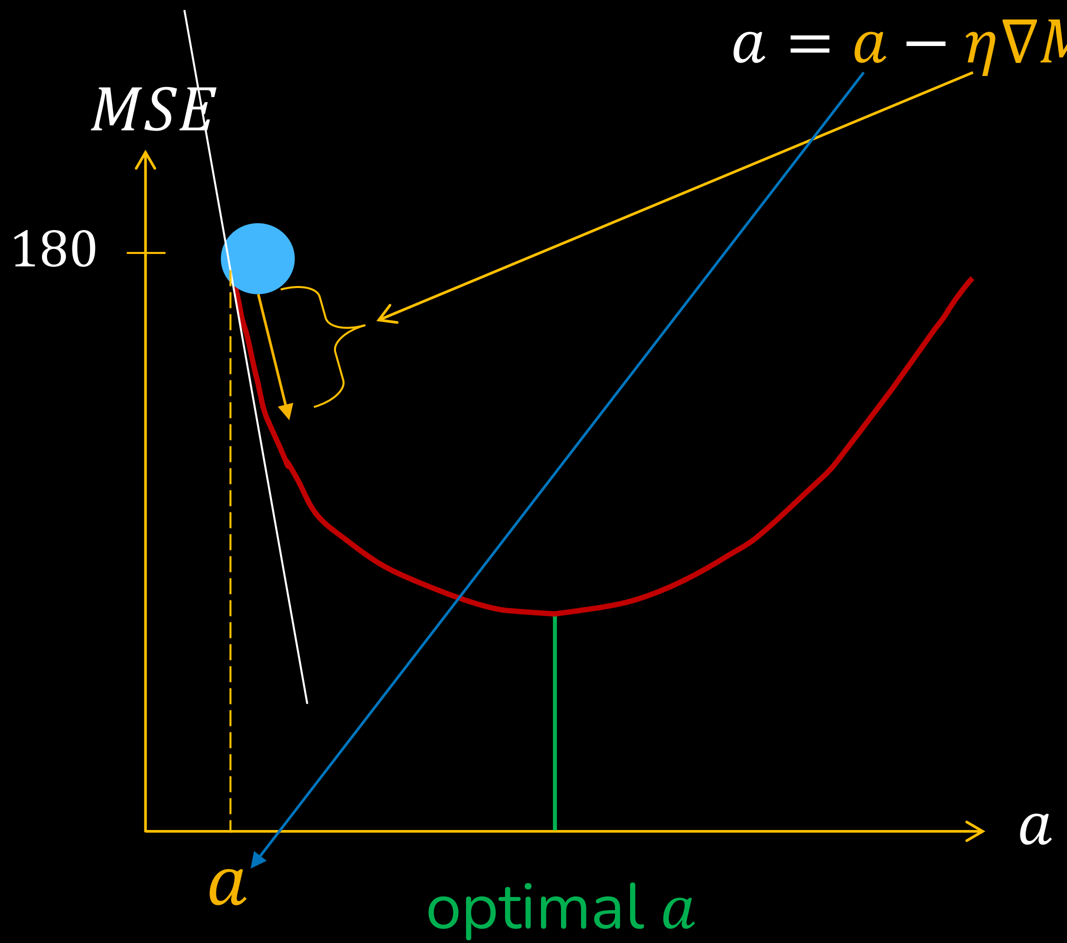
$$a = a - \eta \nabla MSE$$

$$MSE = 180$$



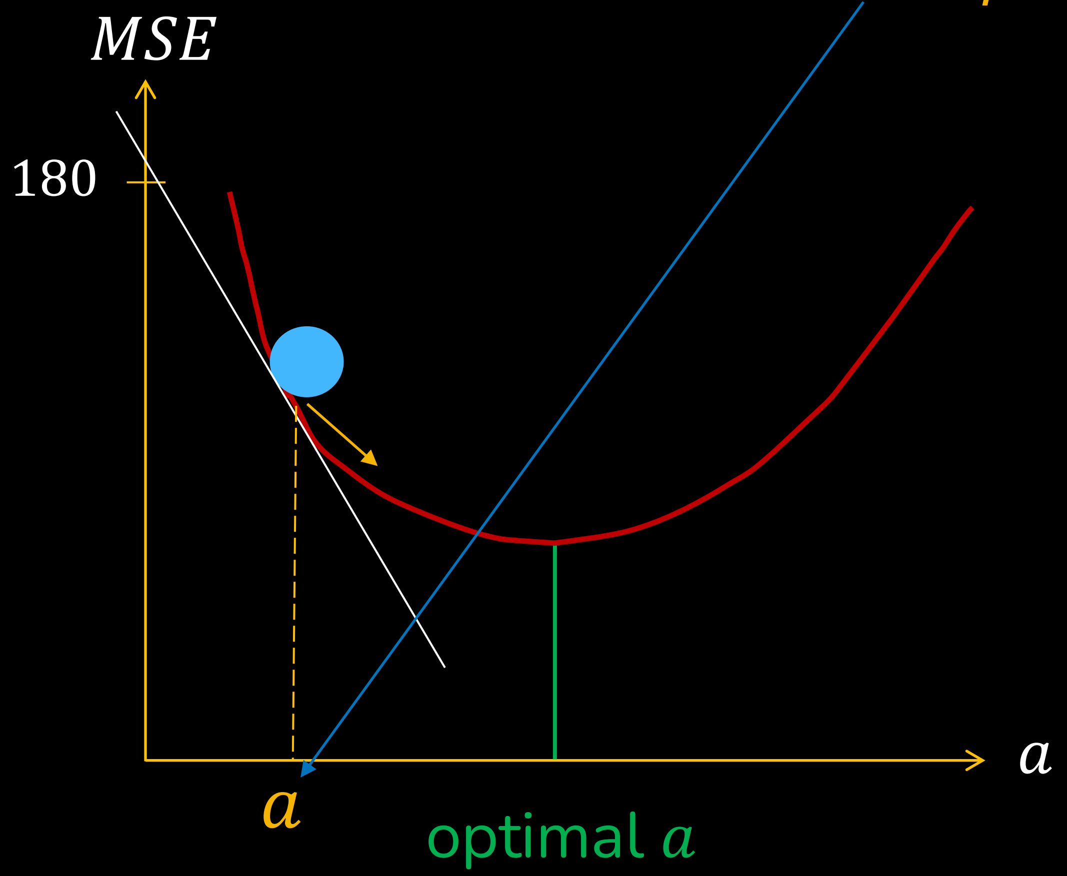
$$a = a - \eta \nabla MSE$$

$MSE = 180$



$$a = a - \eta \nabla MSE$$

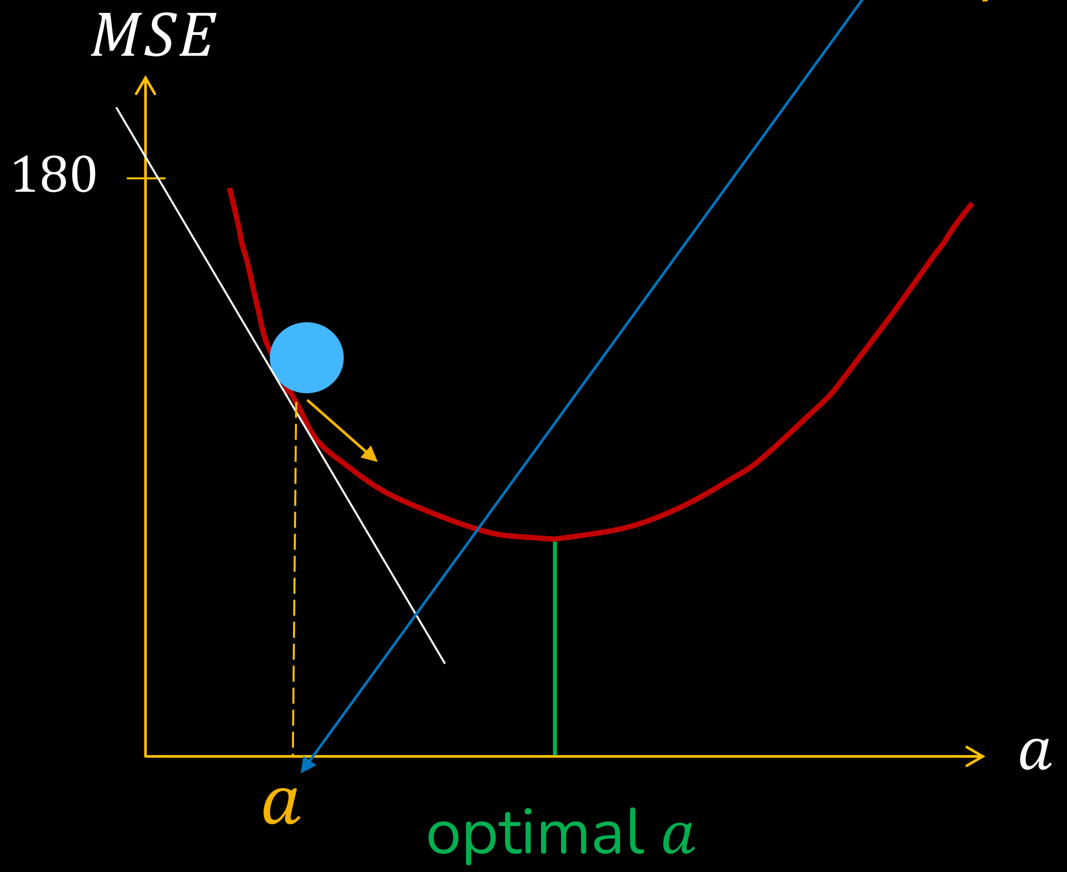
$$MSE = 150$$



$$a = a - \eta \nabla MSE$$

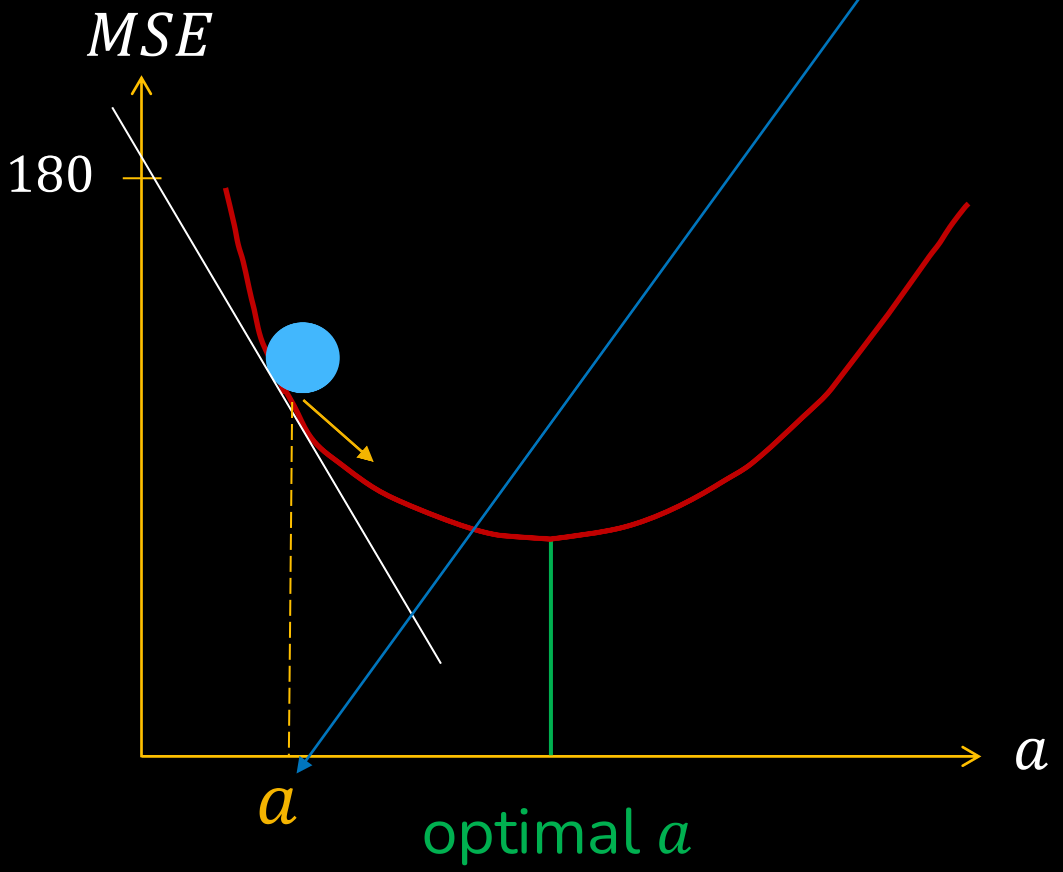
$$MSE = 150$$

$$\nabla MSE = \begin{pmatrix} \frac{\partial MSE(a)}{\partial a} \\ \frac{\partial MSE(b)}{\partial b} \end{pmatrix}$$



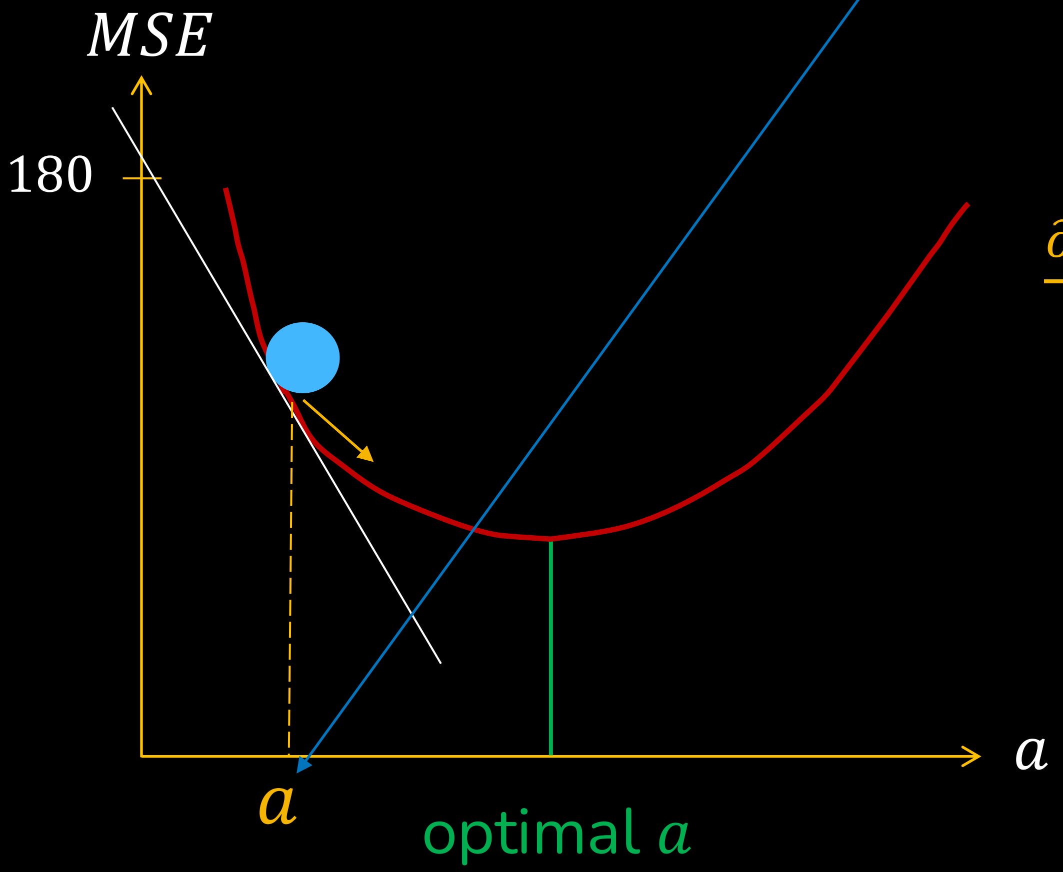
$$a = a - \eta \frac{\partial MSE(a)}{\partial a}$$

$MSE = 150$



$$a = a - \eta \frac{\partial MSE(a)}{\partial a}$$

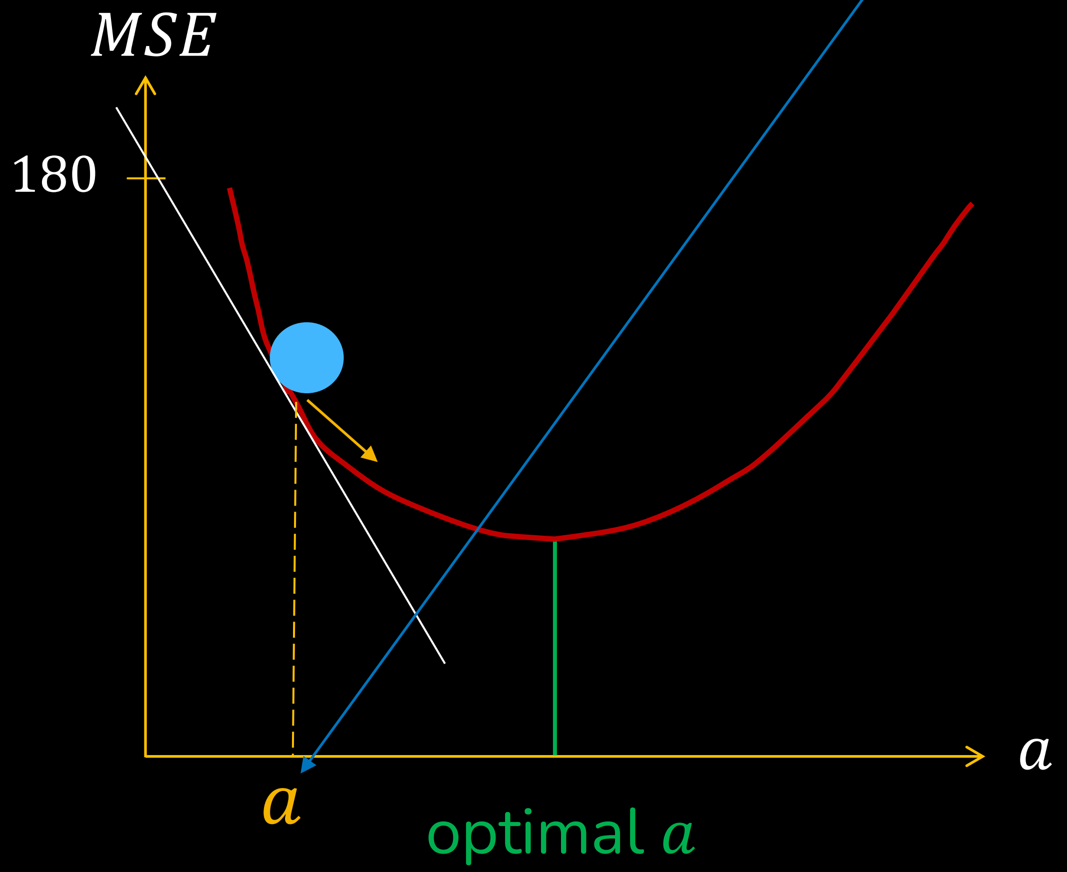
$MSE = 150$



$$\frac{\partial MSE(a)}{\partial a} = \frac{-2}{n} \sum_{i=0}^{n} x_i (y_i - \hat{y}_i)$$

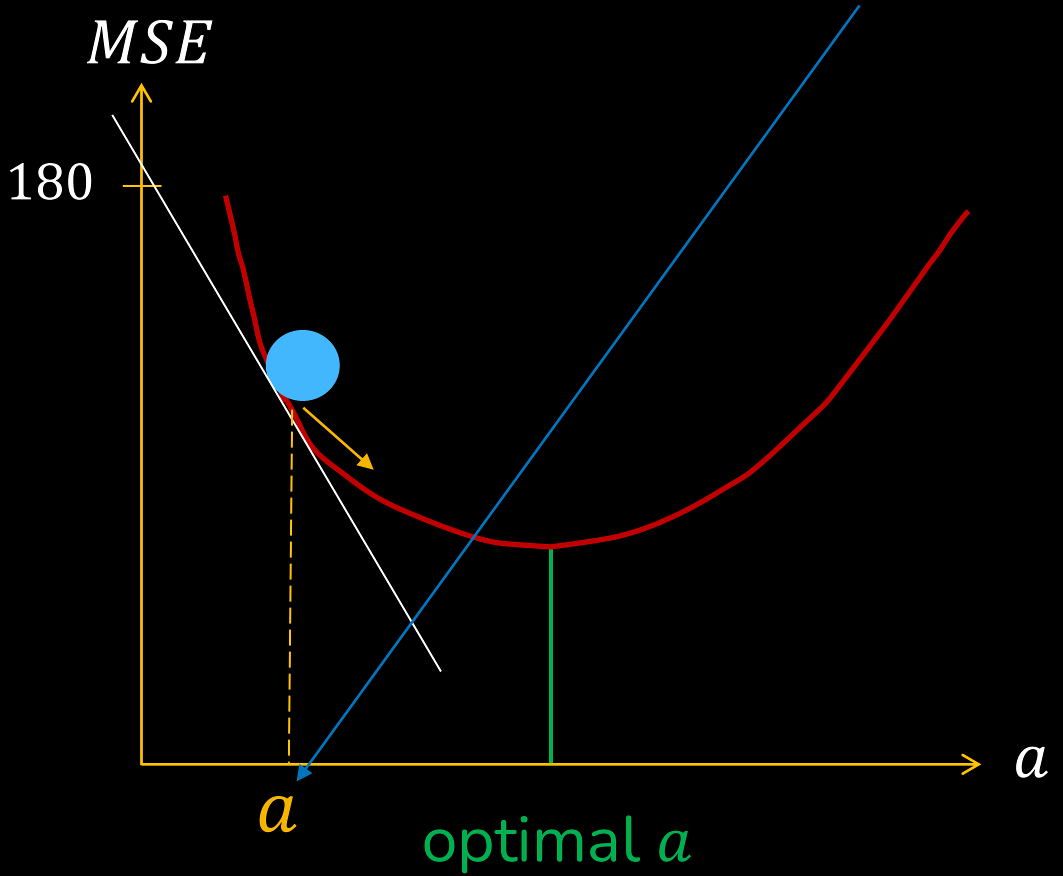
$a = 2 - 0.2$

$MSE = 150$



$a = 1.8$

$MSE = 150$



$$a = 2, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 2.30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$a = 1.8, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 1.8 * 30 + 3$$

$$\hat{y}(x) = 63$$

$$MSE = 180$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$a = 1.8, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 1.8 * 30 + 3$$

$$\hat{y}(x) = 57$$

$$MSE = 180$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$a = 1.8, \qquad b = 3$$

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

Linear regression model

$$\hat{y}(x) = 1.8 * 30 + 3$$

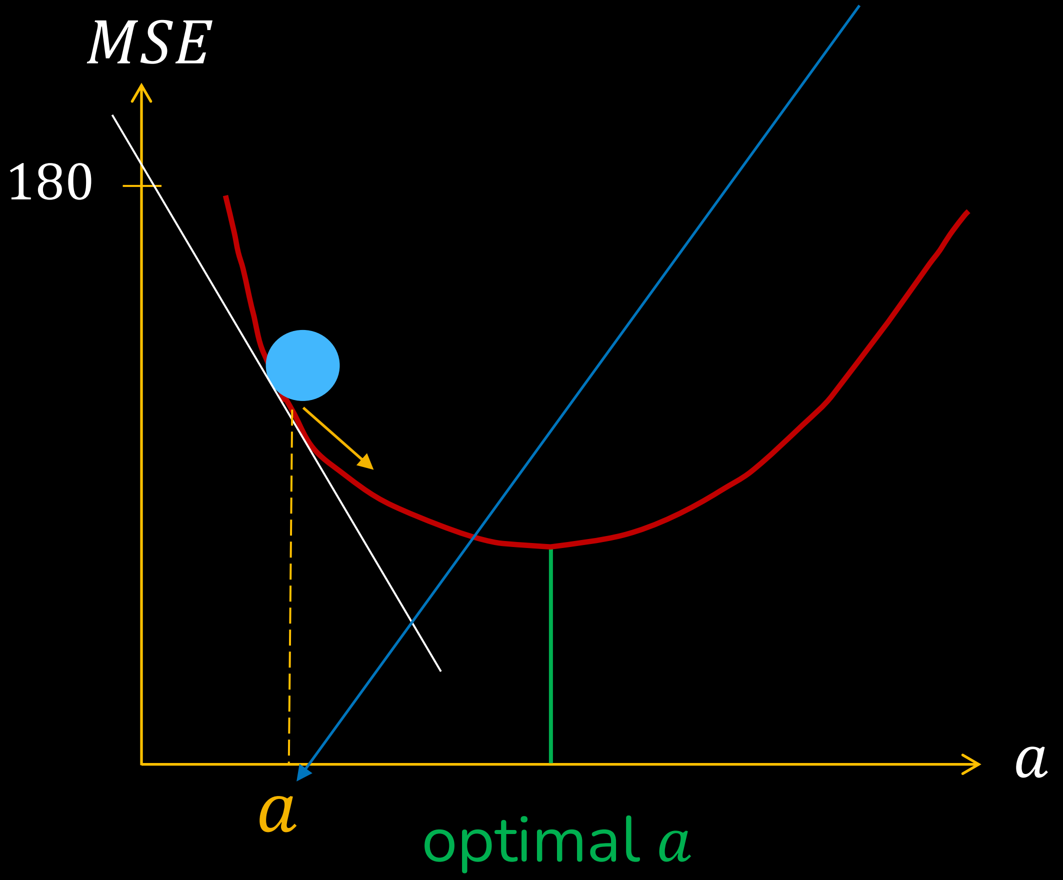
$$\hat{y}(x) = 57$$

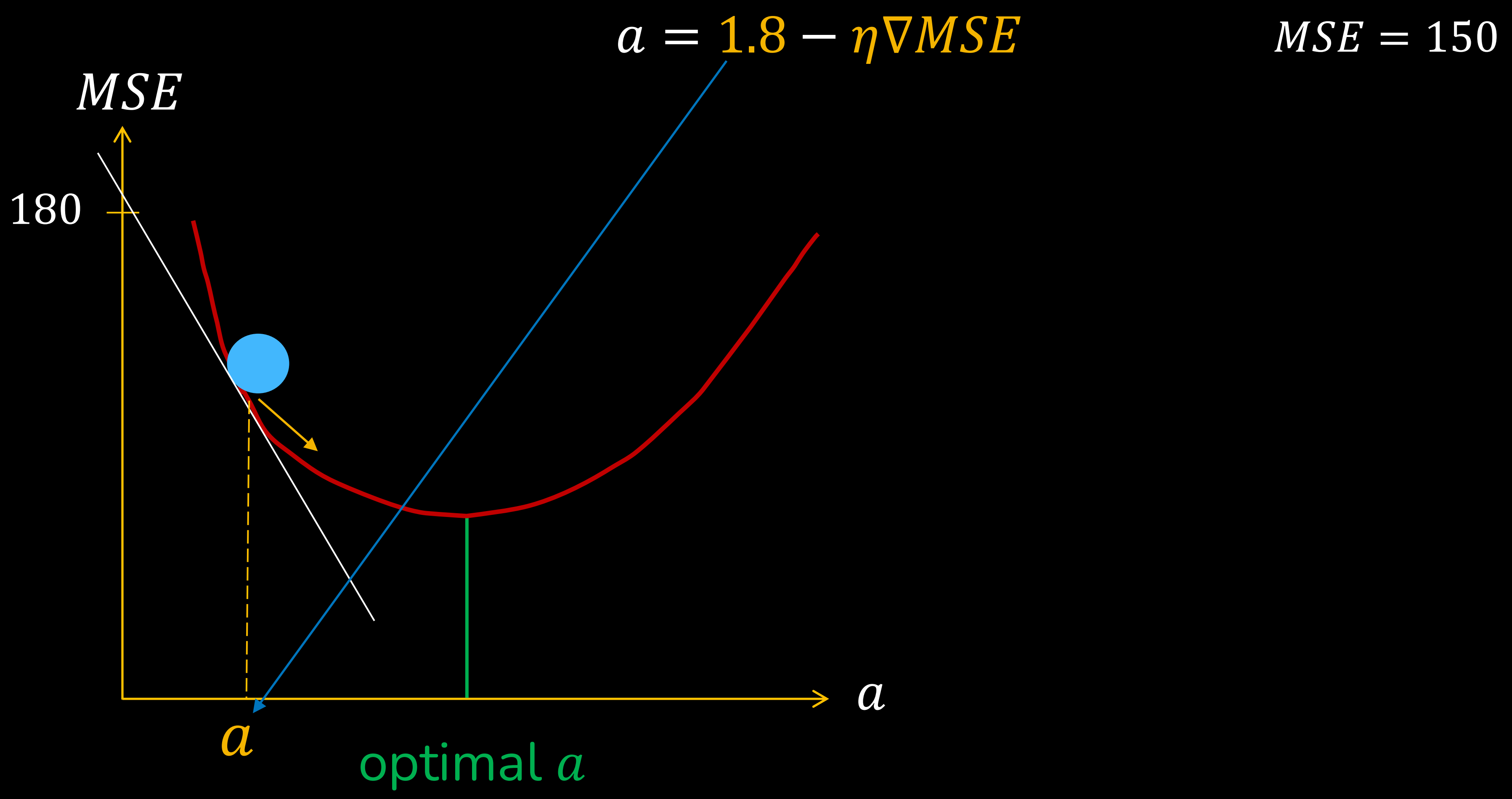
$$MSE = 150$$

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$a = 1.8$

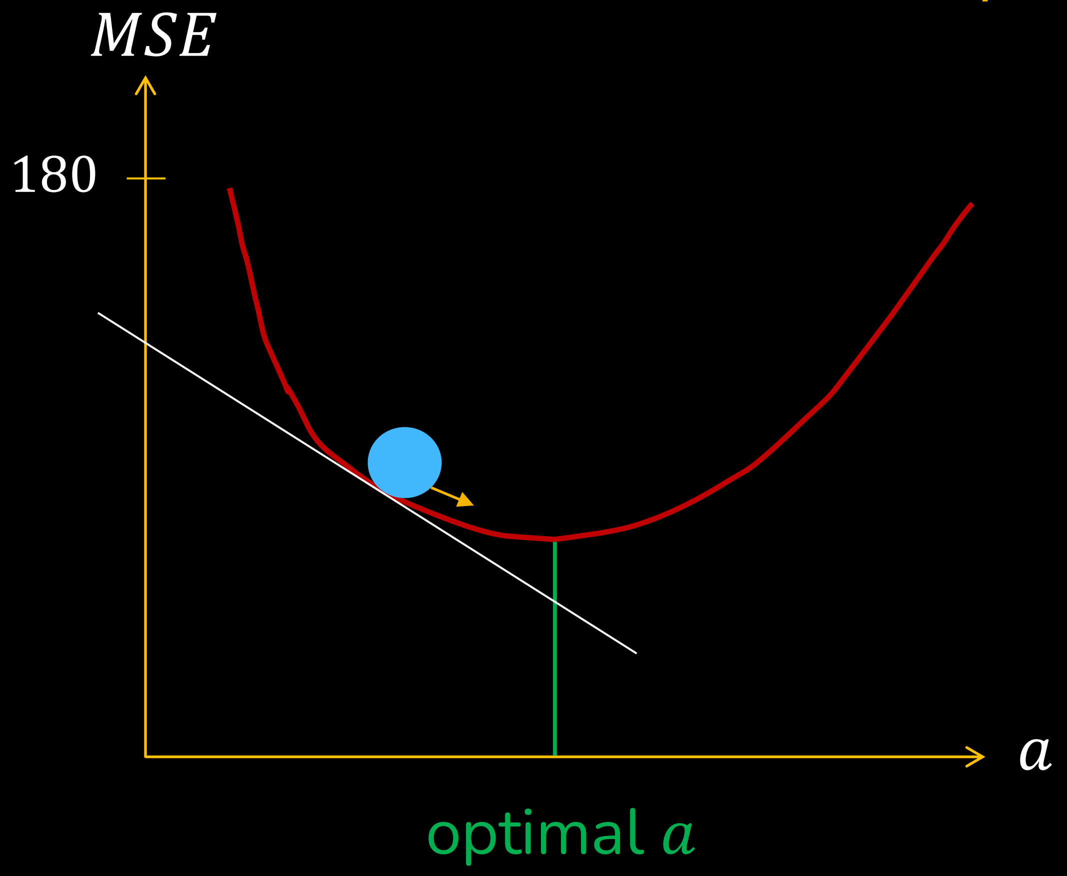
$MSE = 150$




$a = 1.8 - \eta \nabla MSE$
$MSE = 150$
MSE
180
a
a
optimal a

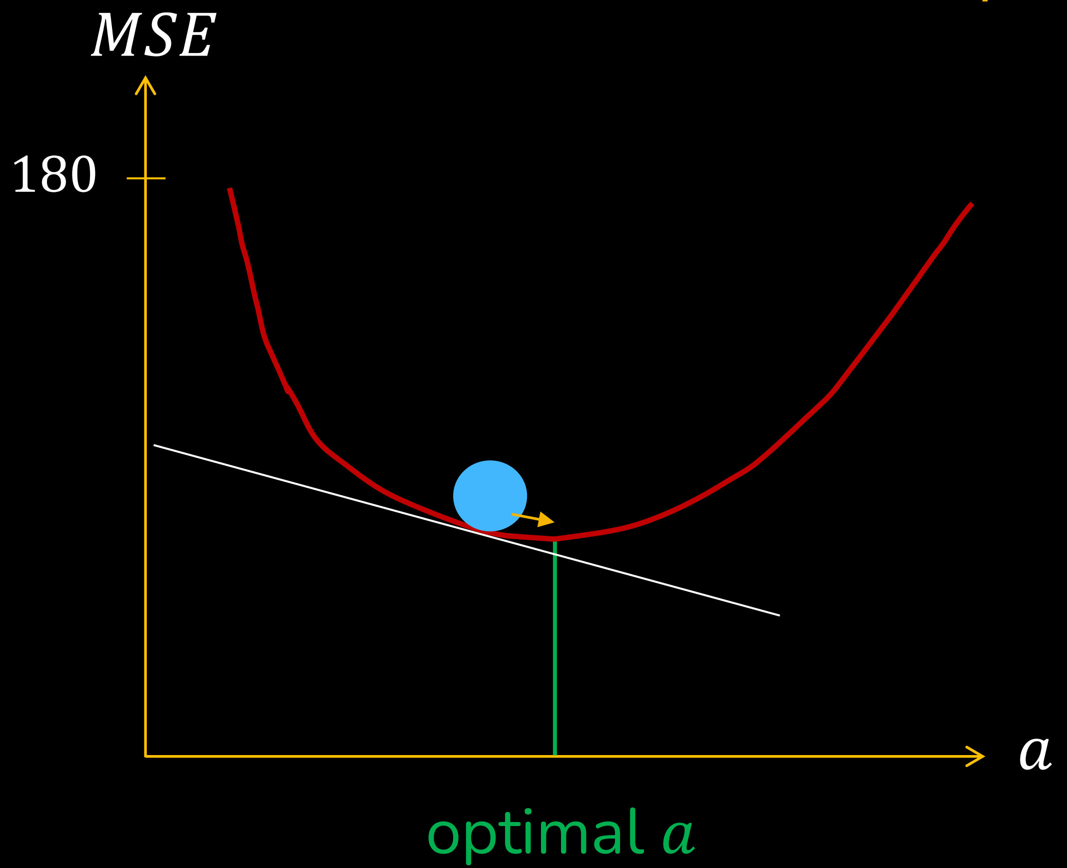
$$a = a - \eta \nabla MSE$$

$MSE = 20$



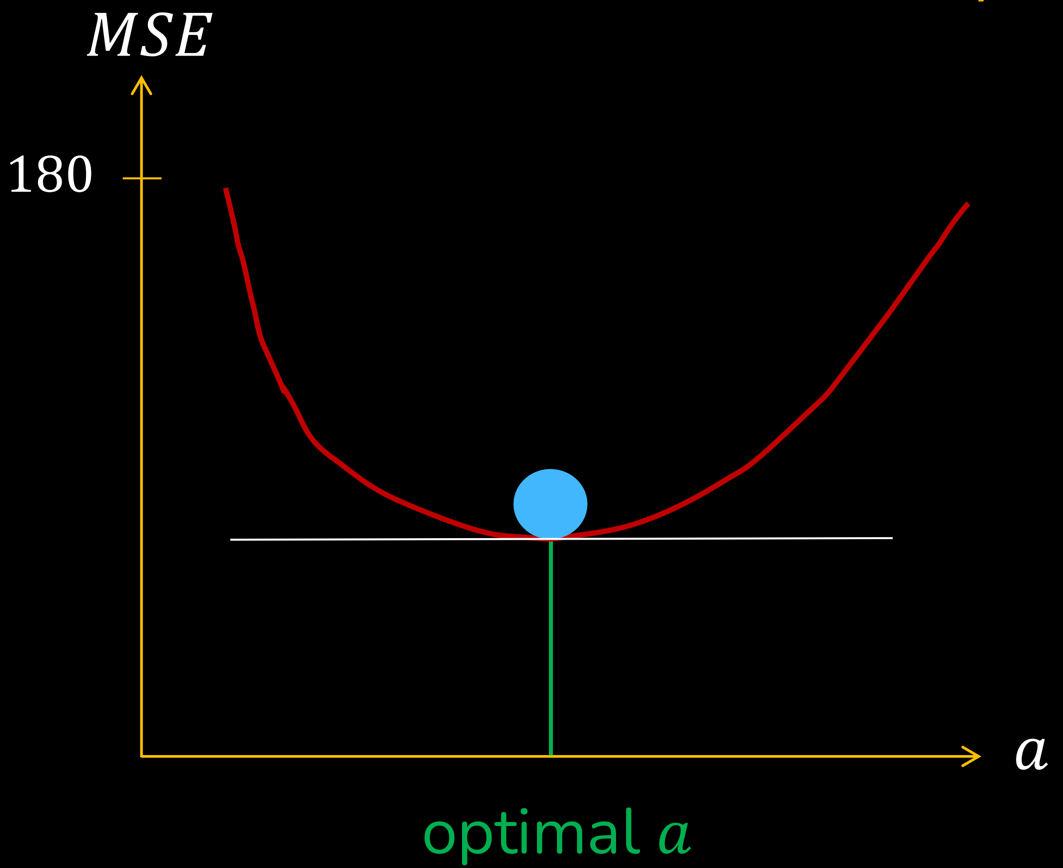
$$a = a - \eta \nabla MSE$$

$$MSE = 10$$



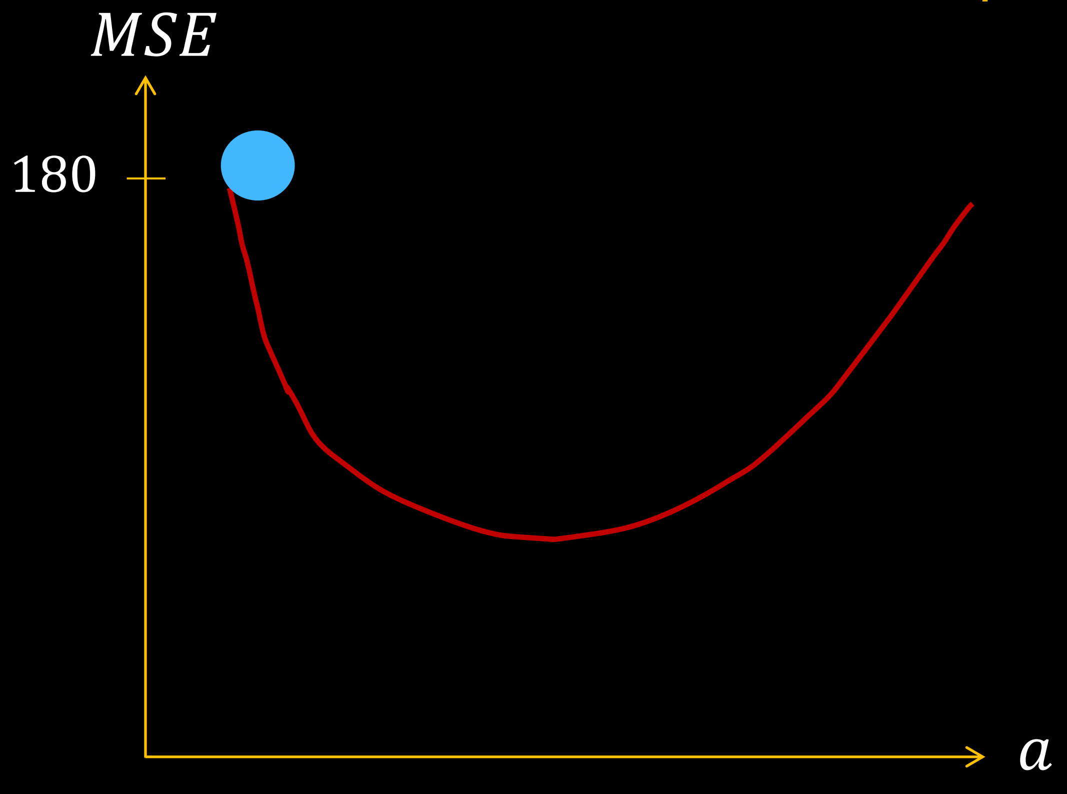
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



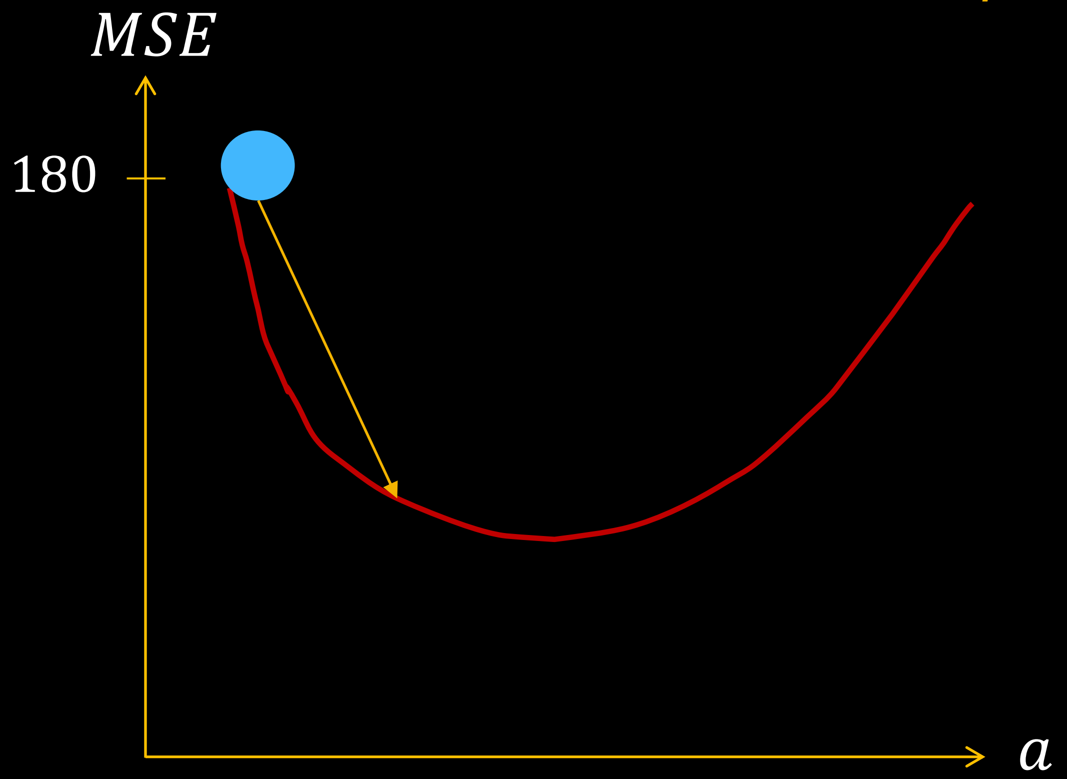
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



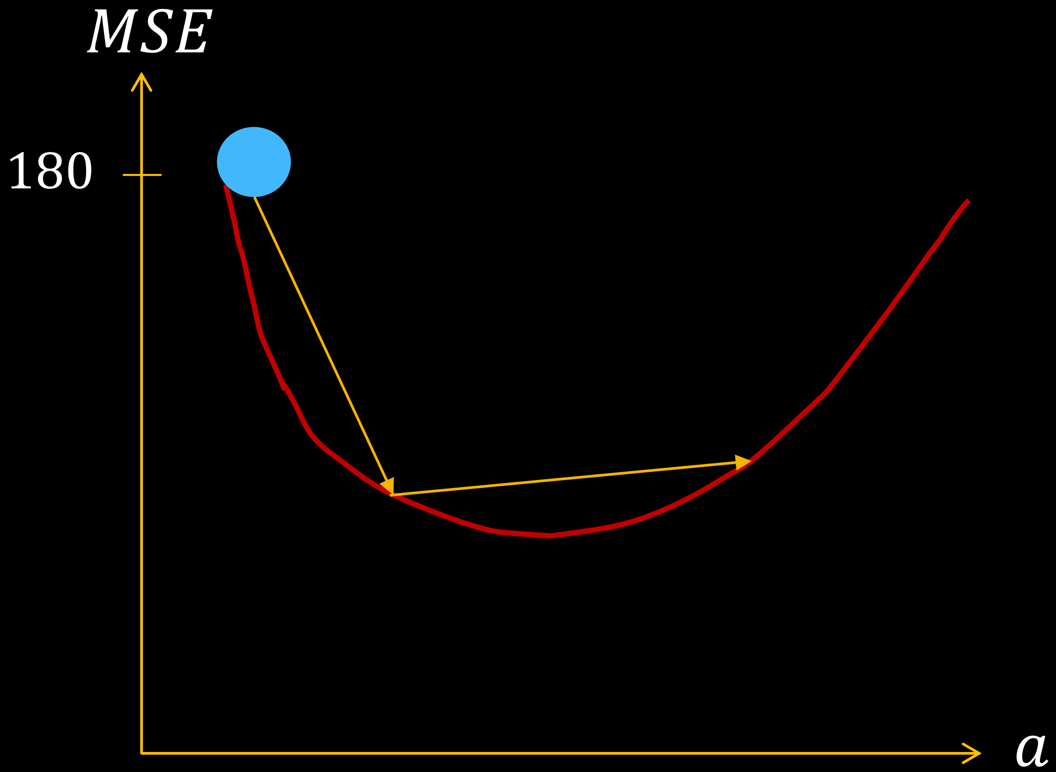
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



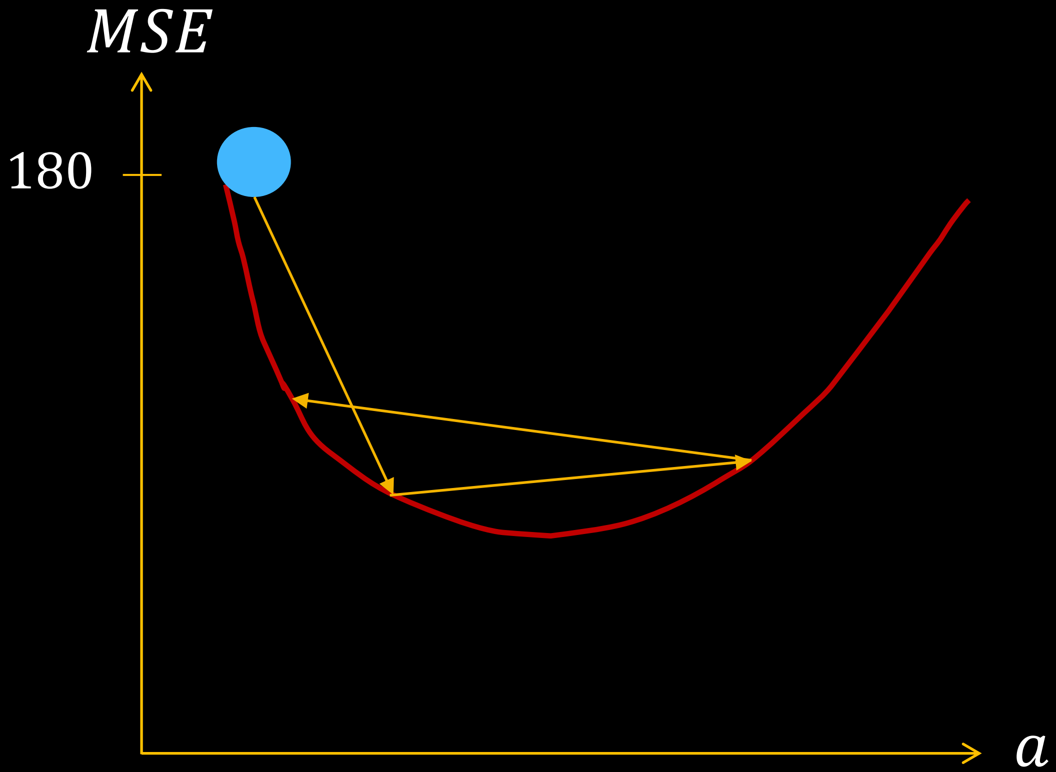
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



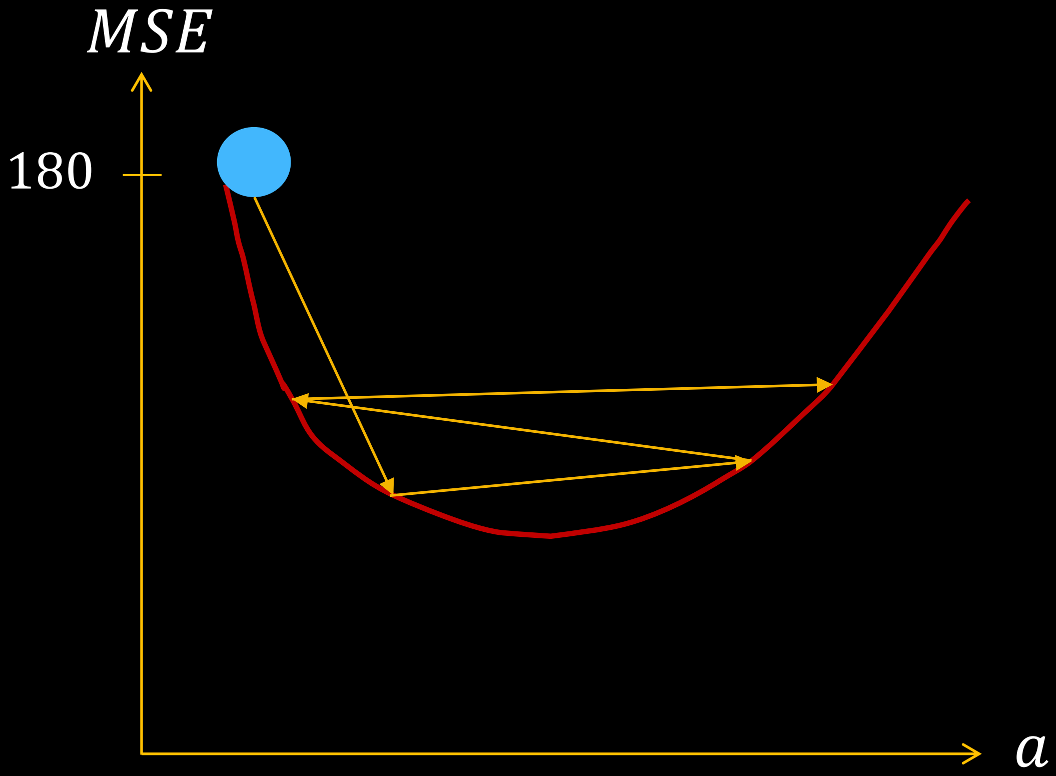
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



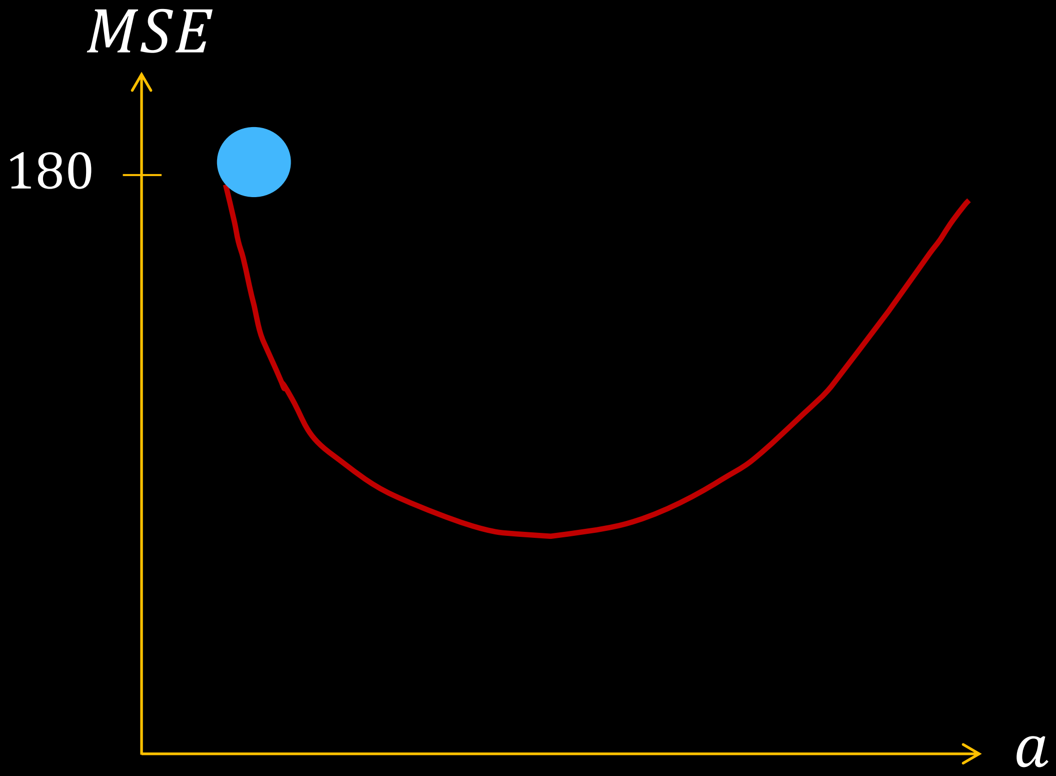
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



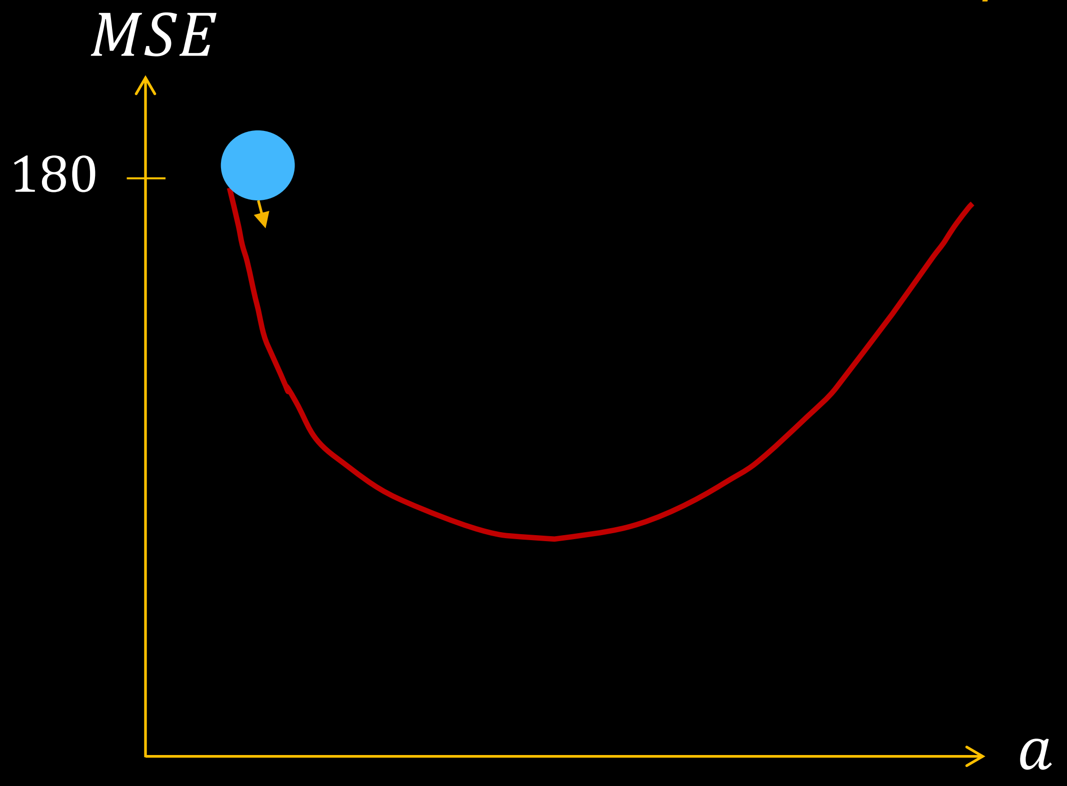
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



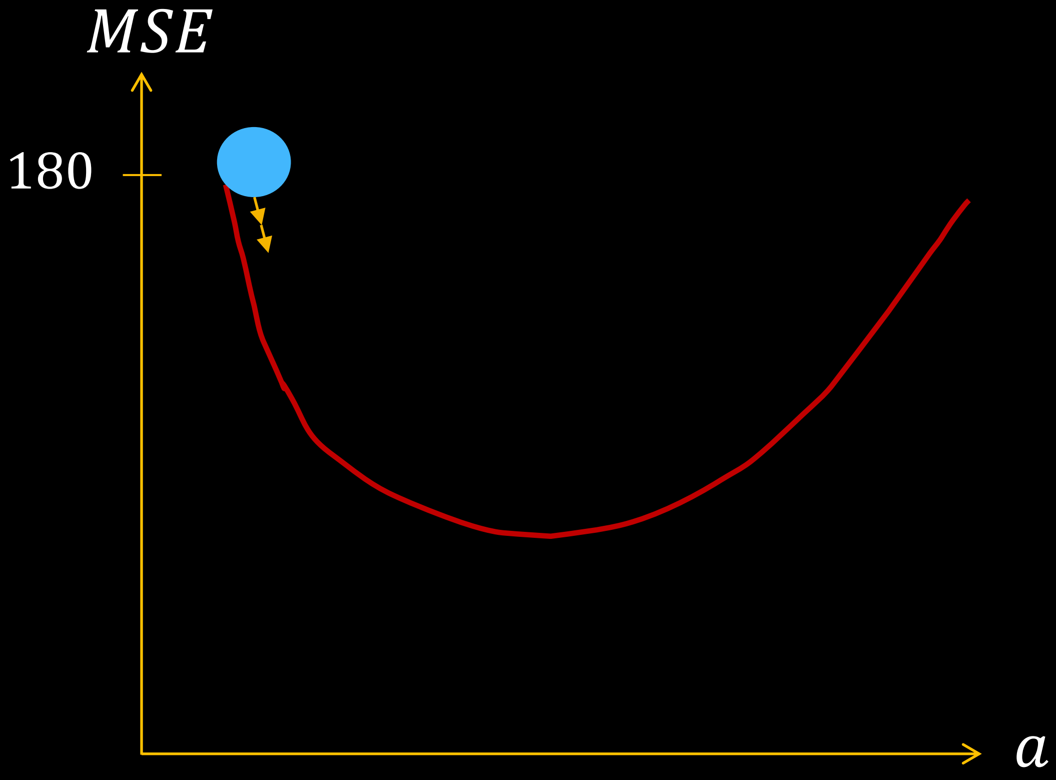
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



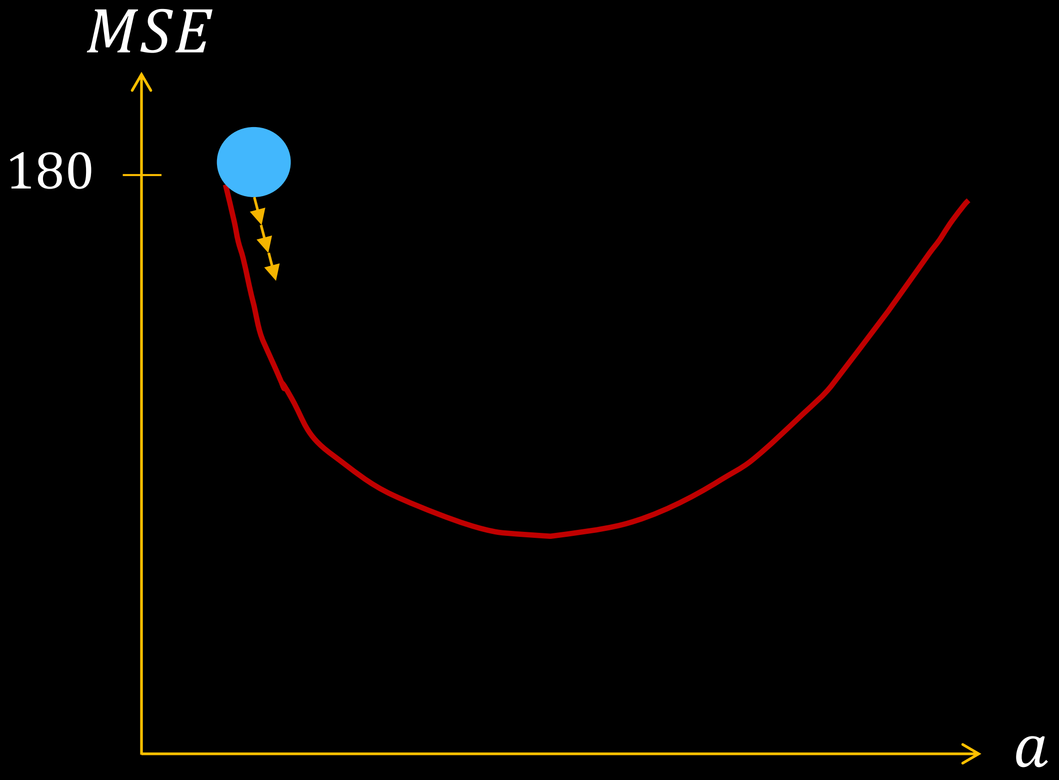
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



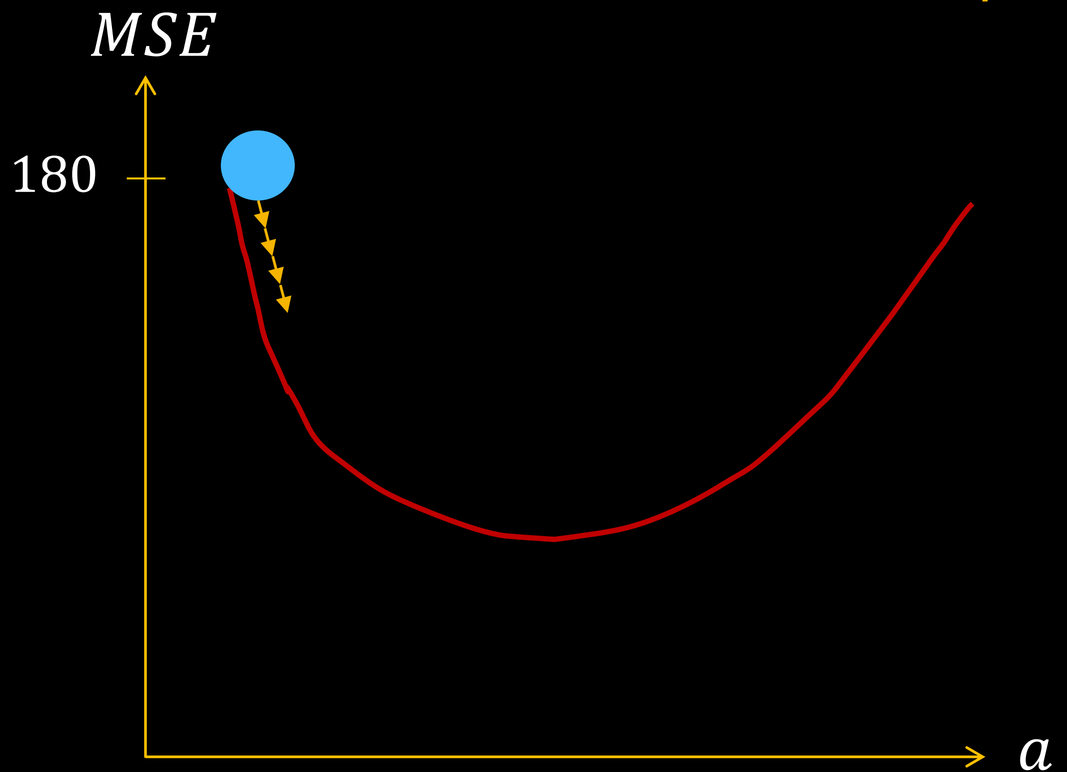
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



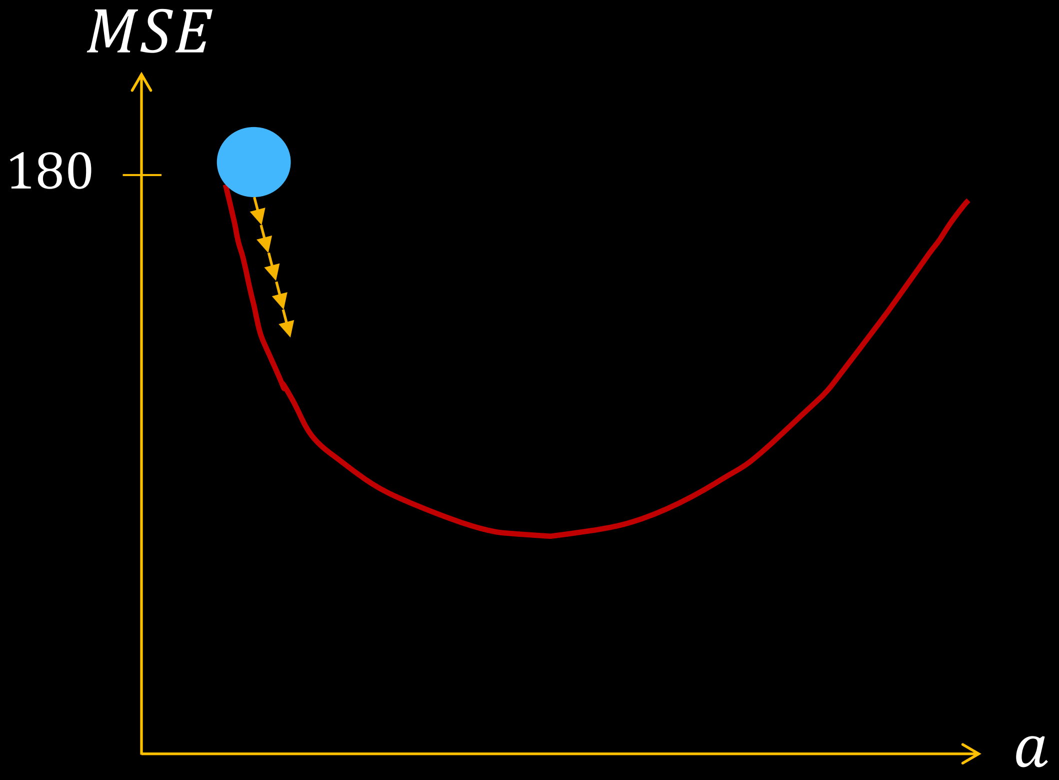
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



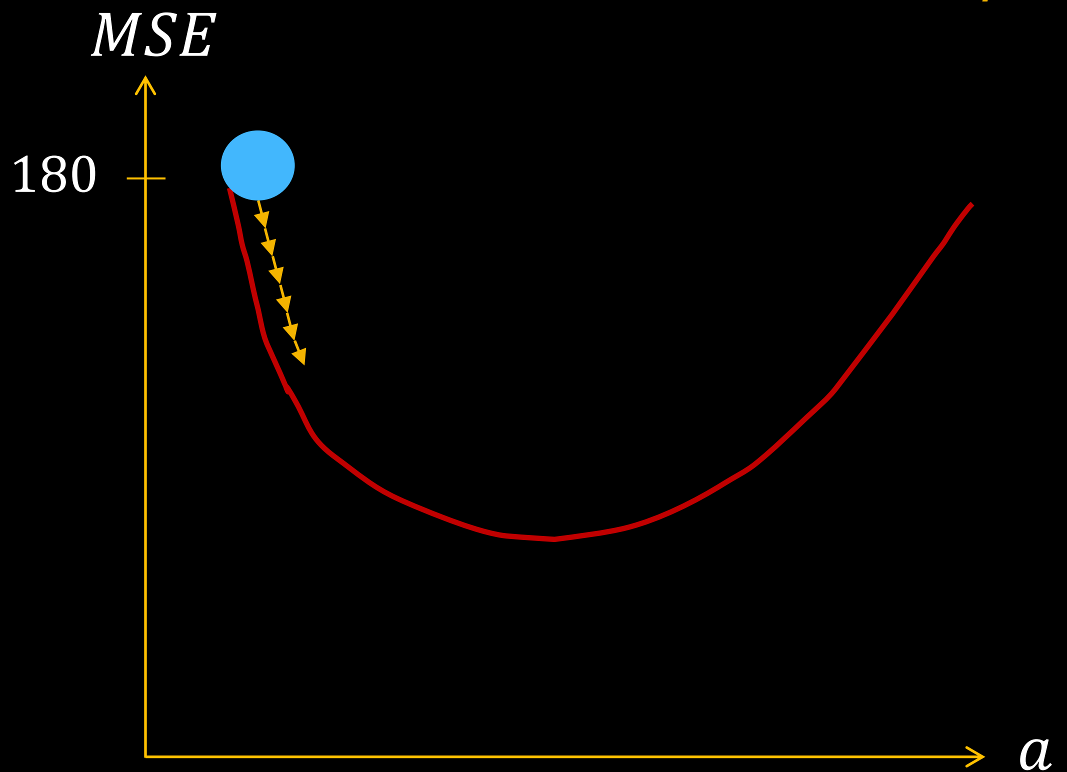
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



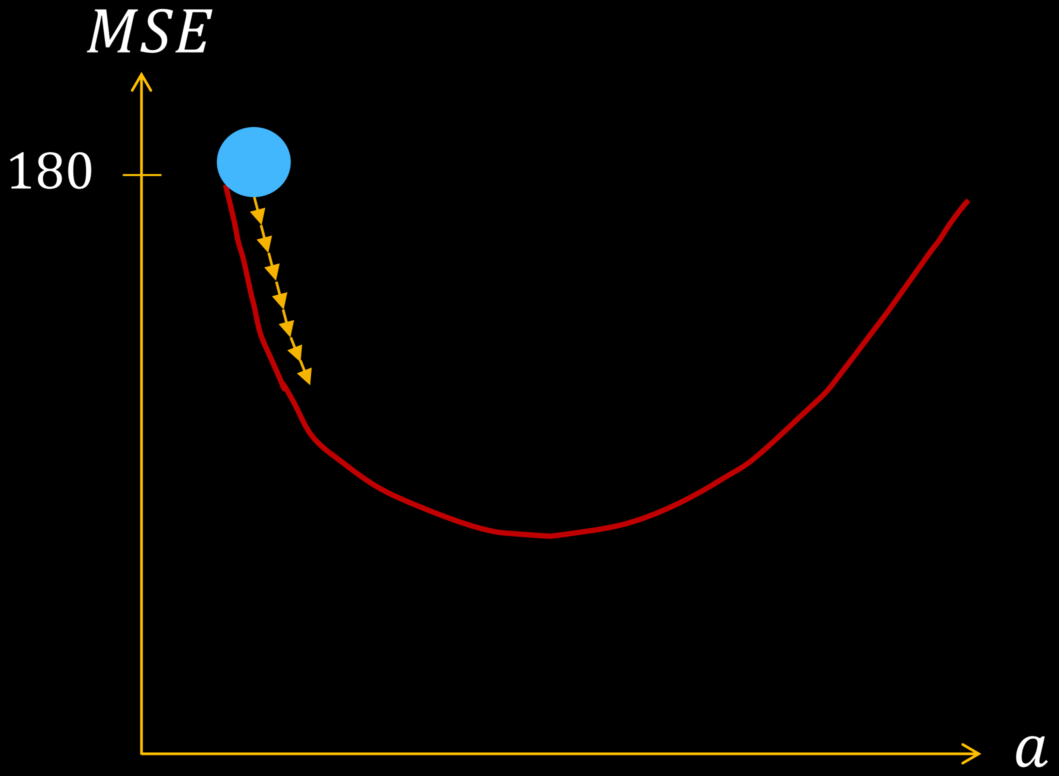
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



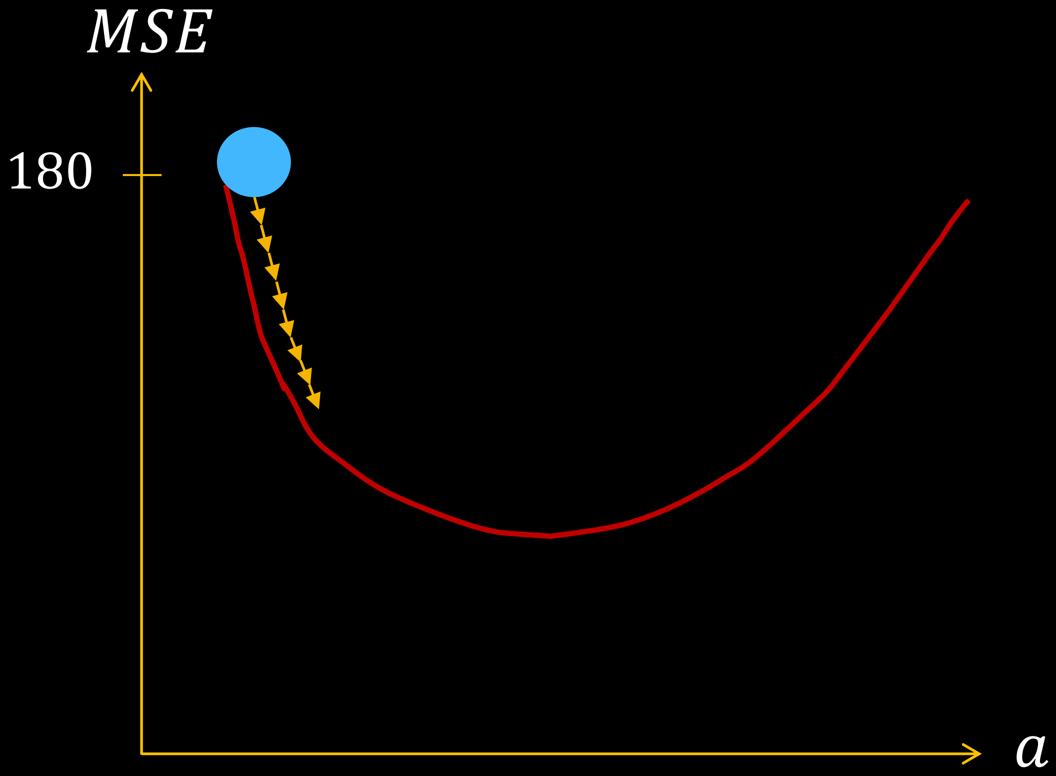
$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$



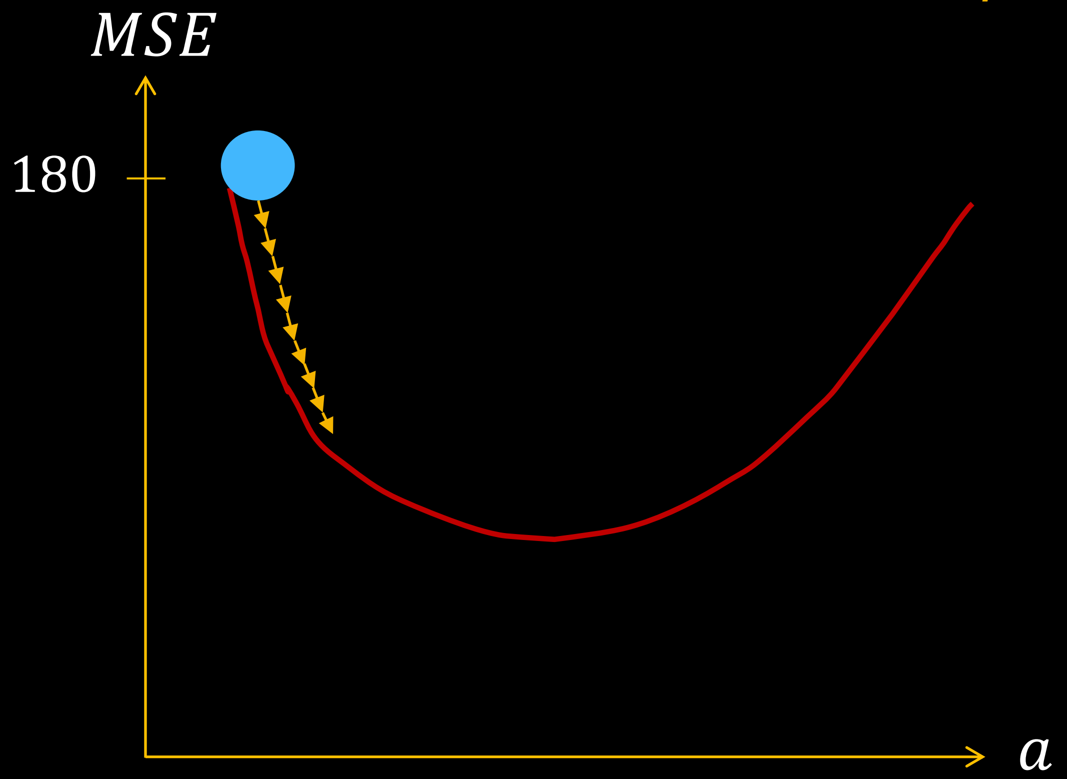
$$a = a - \eta \nabla MSE$$

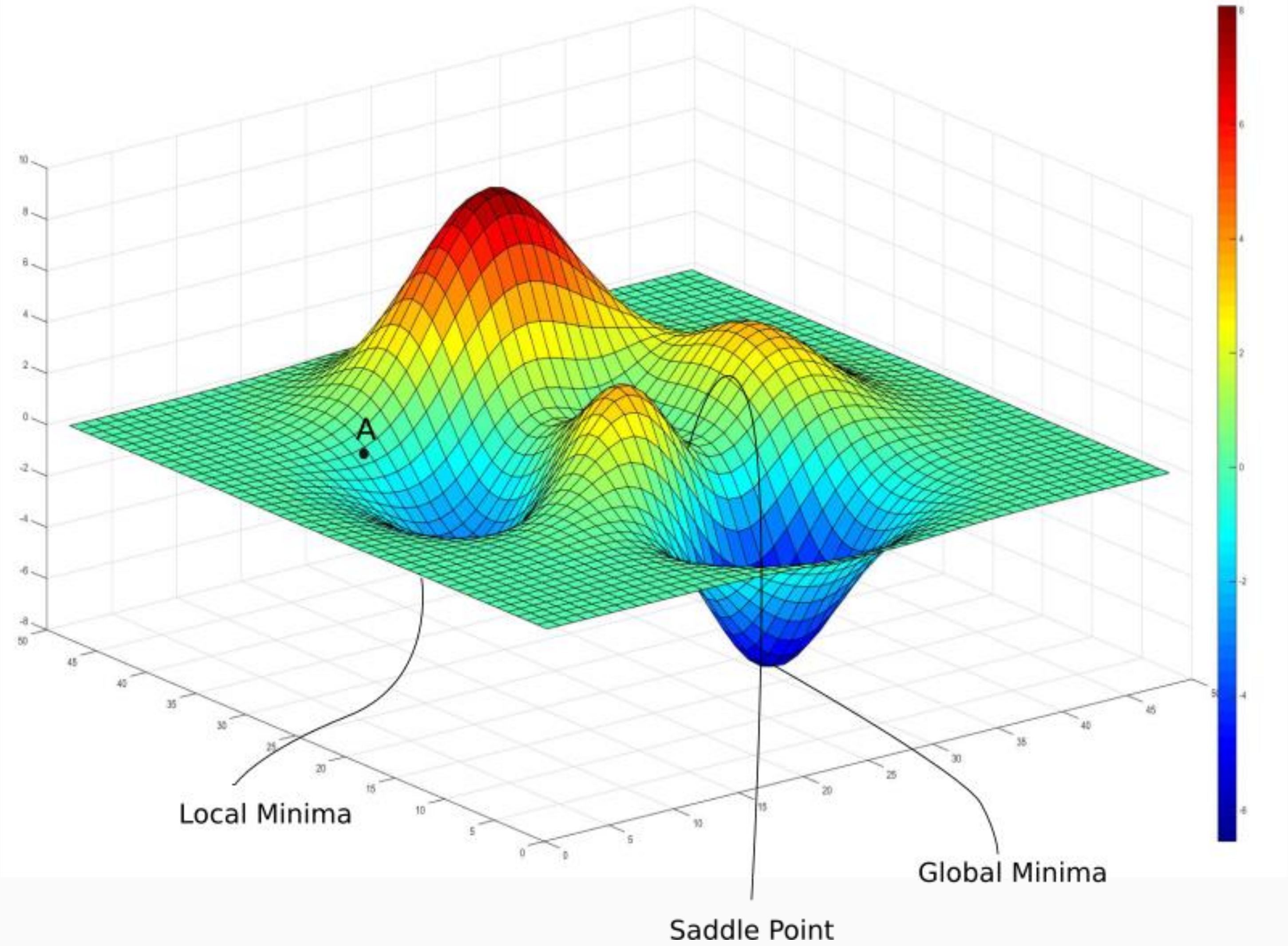
$MSE = 0.001$



$$a = a - \eta \nabla MSE$$

$$MSE = 0.001$$





reference

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$\frac{\partial MSE(a)}{\partial a} = \frac{-2}{n} \sum_{i=0}^{n} x_i (y_i - \hat{y}_i)$$

$$a = a - \eta \frac{\partial MSE(a)}{\partial a}$$

Vanilla gradient descent

temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$\frac{\partial MSE(a)}{\partial a} = \frac{-2}{n} x_i (y_i - \hat{y}_i)$$

$$a = a - \eta \frac{\partial MSE(a)}{\partial a}$$

stochastic gradient descent

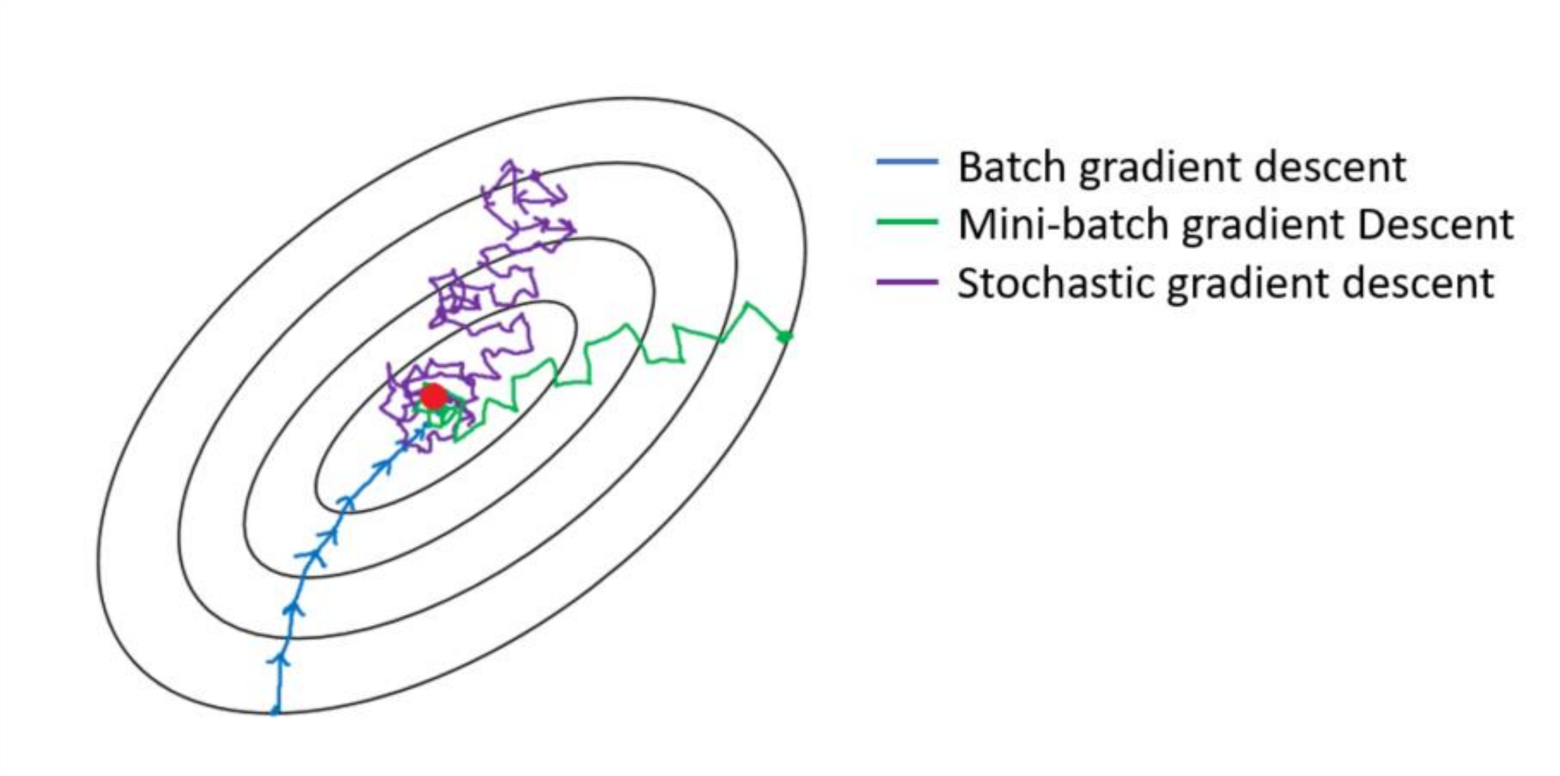
temperature	Ice cream sales
30	56
23	37
0	1
38	87
44	100
…	…
…	…

$$\begin{cases} a = a - \eta \nabla MSE \\ b = b - \eta \nabla MSE \end{cases}$$

$$\frac{\partial MSE(a)}{\partial a} = \frac{-2}{q} \sum_{i=0}^{q} x_i (y_i - \hat{y}_i)$$

$$a = a - \eta \frac{\partial MSE(a)}{\partial a}$$

Mini batch gradient descent


Batch gradient descent
Mini-batch gradient Descent
Stochastic gradient descent

reference

Q&A

END