Multiple Regression

Deriving the Matrix Form

Rahul Singh
rsingh@arrsingh.com

Multiple Regression

Linear Model in
$k + 1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in
$k+1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k+1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in $k + 1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} & . & . & . & \hat{x}_{k0} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} & . & . & . & \hat{x}_{k1} \\ 1 & \hat{x}_{12} & \hat{x}_{22} & \hat{x}_{22} & . & . & . & \hat{x}_{k2} \\ 1 & \hat{x}_{13} & \hat{x}_{23} & \hat{x}_{23} & . & . & . & \hat{x}_{k3} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} & \hat{x}_{3n} & . & . & . & \hat{x}_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k + 1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in
$k + 1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

$$\hat{y}_n = 1 \times \beta_0 + \hat{x}_{1n} \times \beta_1 + \hat{x}_{2n} \times \beta_2 + \hat{x}_{3n} \times \beta_3 + \ldots + \hat{x}_{kn} \times \beta_k$$

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ . \\ . \\ . \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} & . & . & . & \hat{x}_{k0} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} & . & . & . & \hat{x}_{k1} \\ 1 & \hat{x}_{12} & \hat{x}_{22} & \hat{x}_{22} & . & . & . & \hat{x}_{k2} \\ 1 & \hat{x}_{13} & \hat{x}_{23} & \hat{x}_{23} & . & . & . & \hat{x}_{k3} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & \hat{x}_{1n} & \hat{x}_{2n} & \hat{x}_{3n} & . & . & . & \hat{x}_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}$$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k + 1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Multiple Regression

Linear Model in $k + 1$ Dimensions

$$\hat{y} = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2 + \beta_3\hat{x}_3 + \ldots + \beta_k\hat{x}_k$$

$$\hat{Y} = \hat{X}\beta$$

$\hat{Y}$ & β are column vectors. $\hat{X}$ is a matrix

$$\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} \begin{bmatrix} 1 & \hat{x}_{10} & \hat{x}_{20} & \hat{x}_{30} & . & . & . & \hat{x}_{k0} \\ 1 & \hat{x}_{11} & \hat{x}_{21} & \hat{x}_{21} & . & . & . & \hat{x}_{k1} \\ 1 & \hat{x}_{12} & \hat{x}_{22} & \hat{x}_{32} & . & . & . & \hat{x}_{k2} \\ 1 & \hat{x}_{13} & & & . & . & . & \hat{x}_{k3} \\ . & . & & & . & . & . & . \\ . & . & & & . & . & . & . \\ . & . & & & . & . & . & . \\ 1 & \hat{x}_{1n} & x_{2n} & x_{3n} & . & . & . & \hat{x}_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$\hat{Y}$ — $(n+1) \times 1$

$\hat{X}$ — $(n+1) \times (k+1)$

β — $(k+1) \times 1$

- 1 dependent variable $\hat{y}$
- k independent variables $\hat{x}_1, \hat{x}_2, \hat{x}_3 \ldots \hat{x}_k$
- $k + 1$ parameters - $\beta_0, \beta_1, \beta_2, \beta_3 \ldots \beta_k$

Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$

Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

$$\hat{Y} = \hat{X}\beta$$

Matrix form

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$

Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

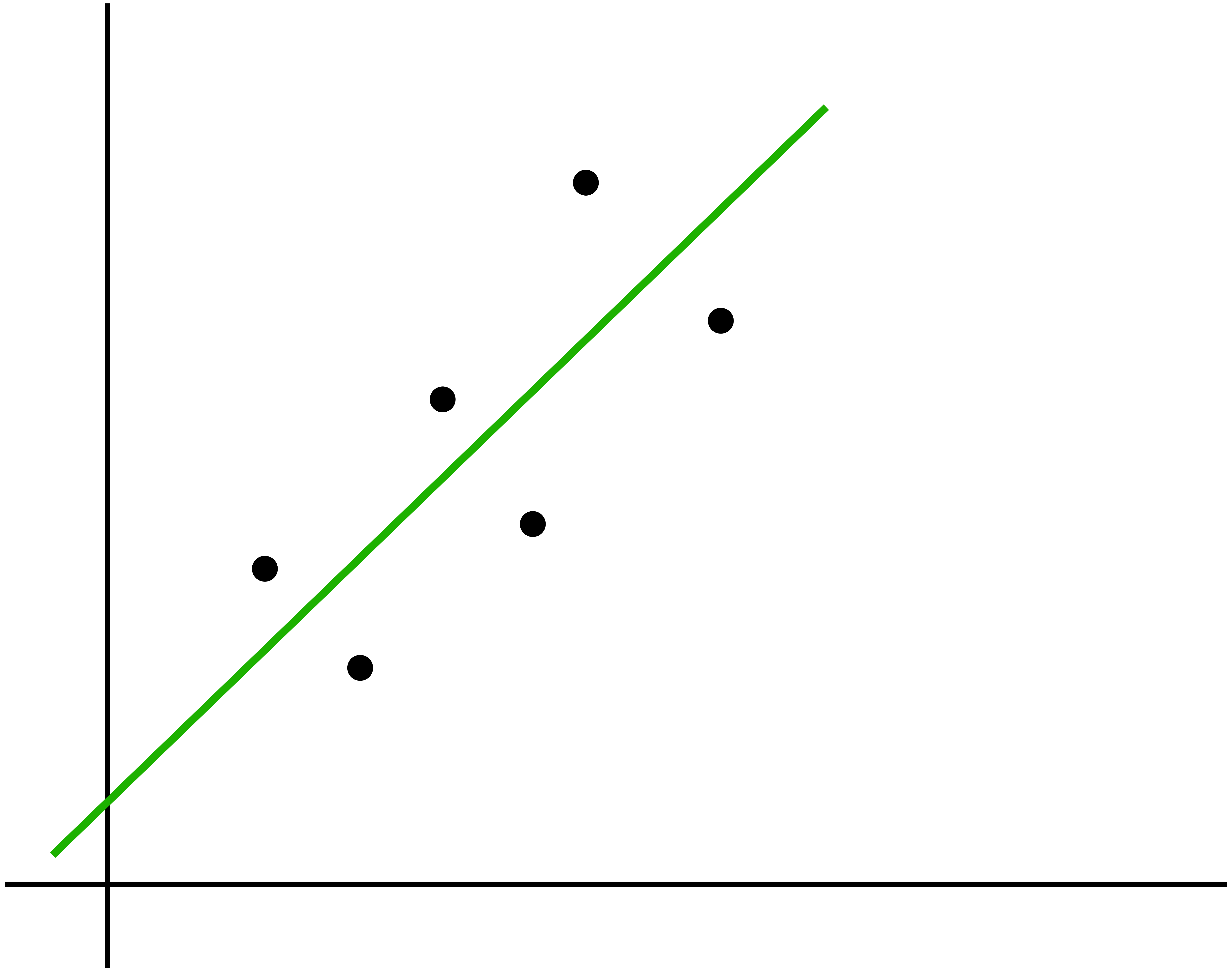
$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

$$\hat{Y} = \hat{X}\beta$$

Matrix form

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$



Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

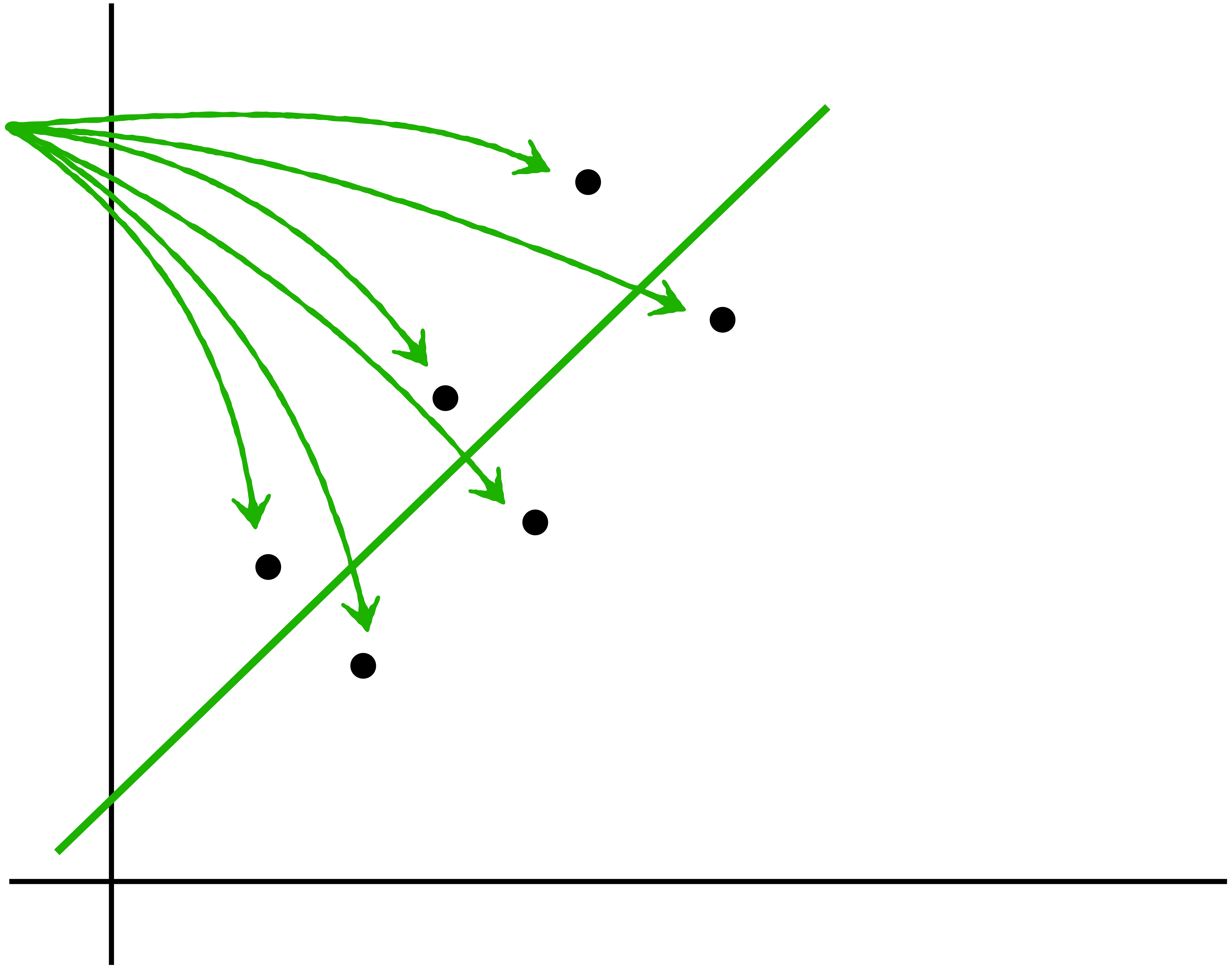
Given a matrix (Y) of observations

$$\hat{Y} = \hat{X}\beta$$

Matrix form

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$



Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

Given a matrix (Y) of observations

$$\hat{Y} = \hat{X}\beta$$

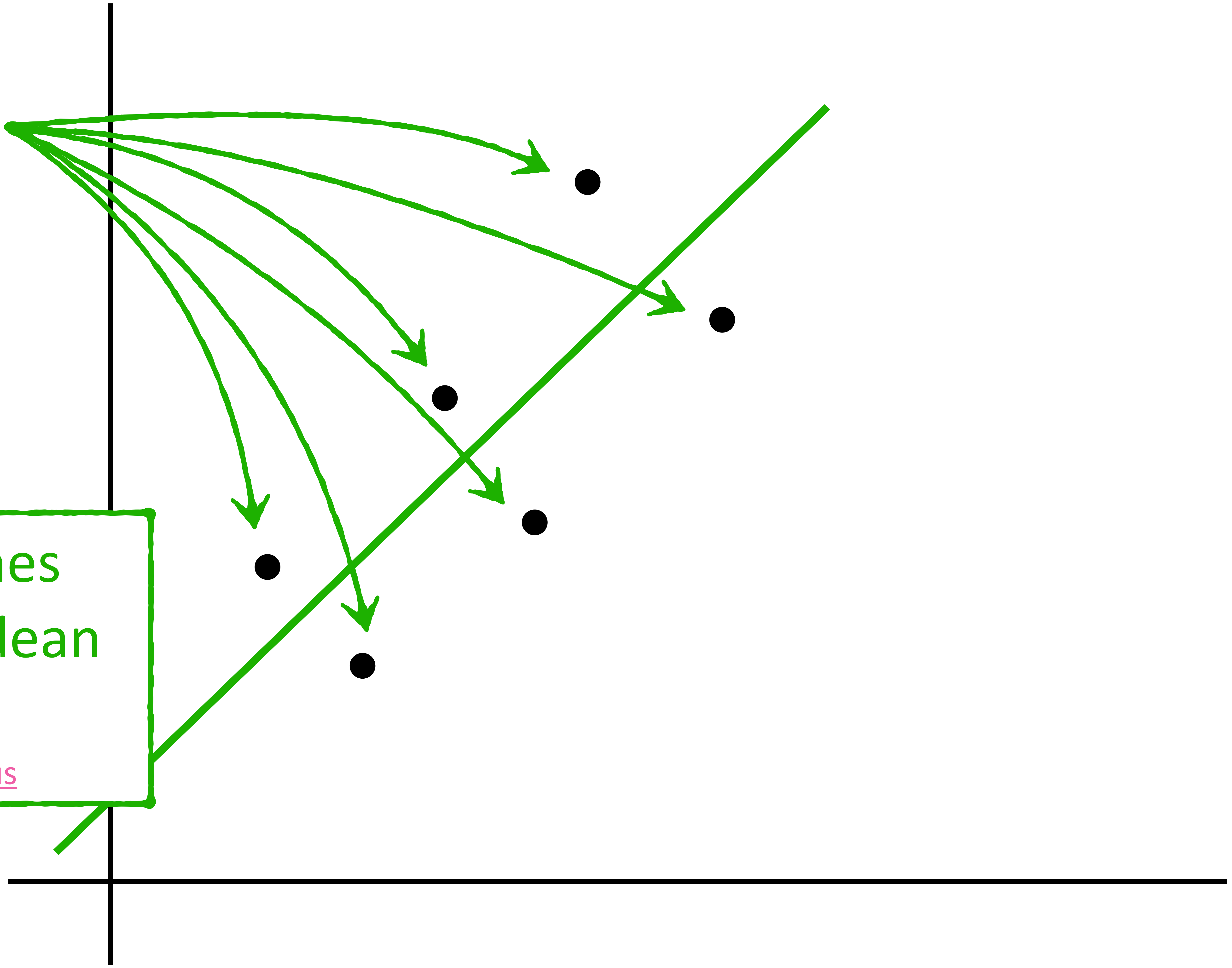
Matrix form

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2$$

The two parallel vertical lines mean that this is the Euclidean Norm of the matrix

See Tutorial on Matrices & Differential Calculus



Data is synthetic and not plotted to scale

Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

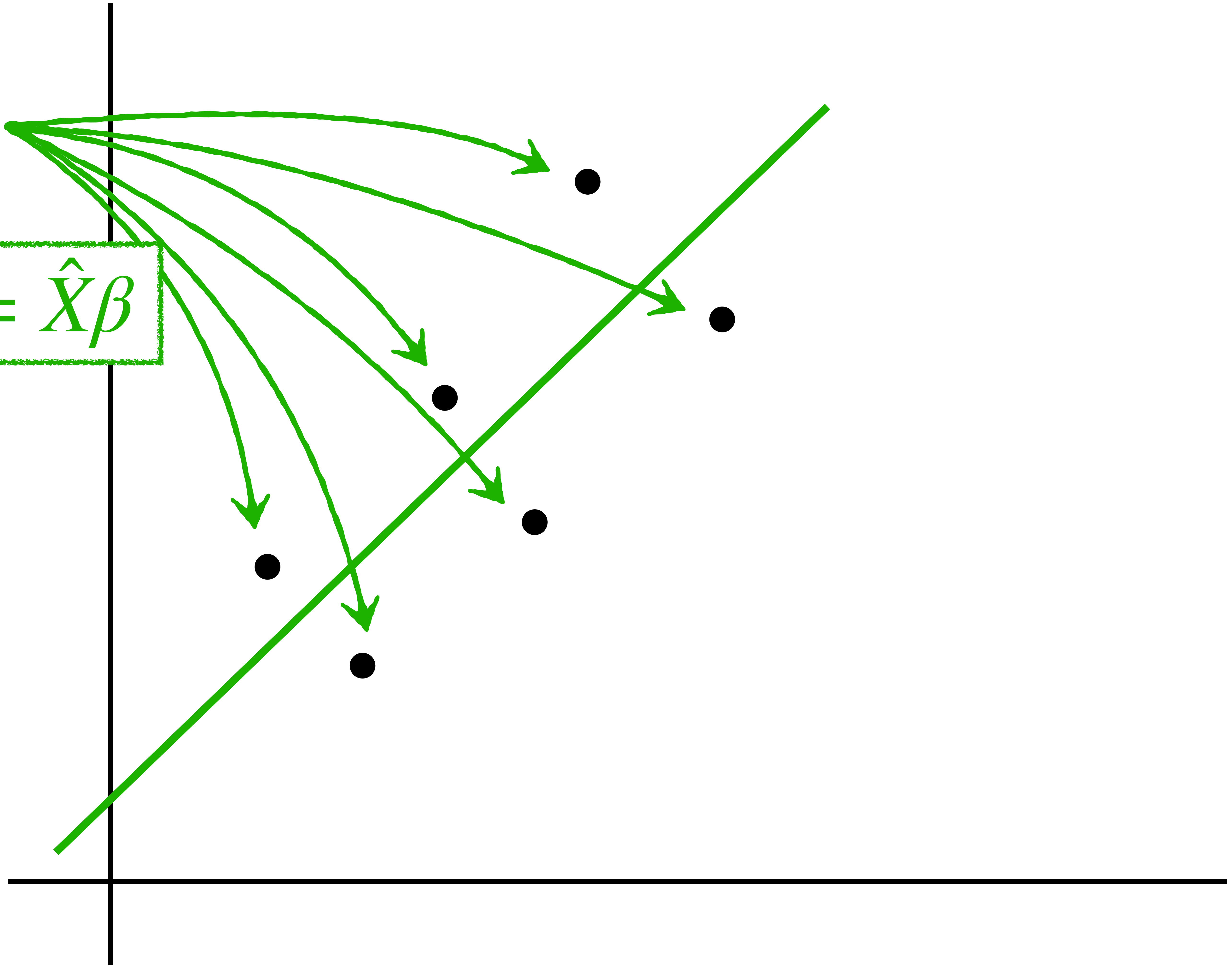
$$\hat{Y} = \hat{X}\beta$$

Given a matrix (Y) of observations

Substituting $\hat{Y} = \hat{X}\beta$

The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2 = \frac{1}{n} \| Y - \hat{X}\beta \|^2$$



Data is synthetic and not plotted to scale

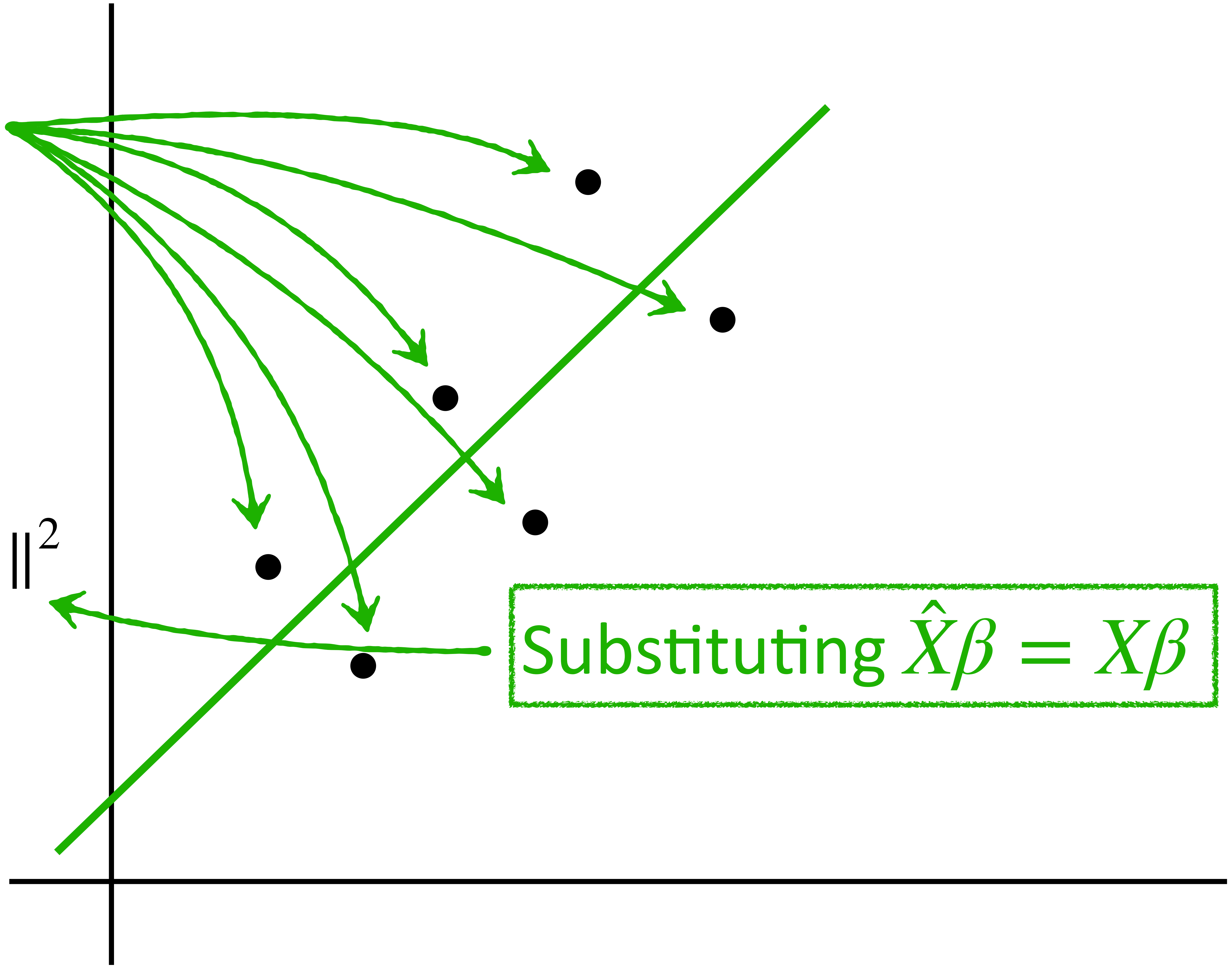
Simple Linear Regression

Linear Model in two Dimensions

$$\hat{y} = \beta_0 + \beta_1 \hat{x}_1$$

$$\hat{Y} = \hat{X}\beta$$

Given a matrix (Y) of observations



The **Mean Squared Error (MSE)**

$$\frac{1}{n} \| Y - \hat{Y} \|^2 = \frac{1}{n} \| Y - \hat{X}\beta \|^2 = \frac{1}{n} \| Y - X\beta \|^2$$

Substituting $\hat{X}\beta = X\beta$

Data is synthetic and not plotted to scale

Linear Model in $k+1$ Dimensions

$$\hat{Y} = \hat{X}\beta$$

The **Mean Squared Error (MSE):**

$$\frac{1}{n} \| Y - X\beta \|^2$$

Linear Model in $k+1$ Dimensions

$$\hat{Y} = \hat{X}\beta$$

The **Mean Squared Error (MSE):**

$$\frac{1}{n} \| Y - X\beta \|^2$$

The Problem Statement:

Multiple Regression: Compute the matrix β such that the Mean Squared Error (MSE) is minimized.

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n}\| Y - X\beta \|^2$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n} \| Y - X\beta \|^2$$

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n} \| Y - X\beta \|^2$$

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

Euclidean norm of a matrix:

$$\| A \| = \sqrt{A^T A}$$

See the Tutorial on Vectors & Matrices

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n} \| Y - X\beta \|^2$$

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

$$\Rightarrow \frac{1}{n} \frac{\partial}{\partial \beta} \left(\sqrt{(Y - X\beta)^T (Y - X\beta)} \right)^2 = 0$$

Euclidean norm of a matrix:

$$\| A \| = \sqrt{A^T A}$$

See the Tutorial on Vectors & Matrices

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n} \| Y - X\beta \|^2$$

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

$$\Rightarrow \frac{1}{n} \frac{\partial}{\partial \beta} \left(\sqrt{(Y - X\beta)^T (Y - X\beta)} \right)^2 = 0$$

Euclidean norm of a matrix:

$$\| A \| = \sqrt{A^T A}$$

See the Tutorial on Vectors & Matrices

$$\Rightarrow \frac{\partial}{\partial \beta} (Y - X\beta)^T (Y - X\beta) = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

To derive value of the matrix β we calculate the partial derivative of the Mean Squared Error (MSE) w.r.t β and solve for β

$$\frac{1}{n} \| Y - X\beta \|^2$$

$$\frac{\partial}{\partial \beta} \frac{1}{n} \| Y - X\beta \|^2 = 0$$

$$\Rightarrow \frac{1}{n} \frac{\partial}{\partial \beta} \left(\sqrt{(Y - X\beta)^T (Y - X\beta)} \right)^2 = 0$$

$$\Rightarrow \frac{\partial}{\partial \beta} \left(\sqrt{(Y - X\beta)^T (Y - X\beta)} \right)^2 = 0$$

$$\Rightarrow \frac{\partial}{\partial \beta} (Y - X\beta)^T (Y - X\beta) = 0$$

Euclidean norm of a matrix:

$$\| A \| = \sqrt{A^T A}$$

See the Tutorial on Vectors & Matrices

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial \beta}(Y - X\beta)^T(Y - X\beta) = 0$$

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - \frac{\partial}{\partial \beta}Y^TX\beta - \frac{\partial}{\partial \beta}\beta^TX^TY + \frac{\partial}{\partial \beta}\beta^TX^TX\beta = 0$$

Expanding to the derivatives of the individual terms

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial \beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - \frac{\partial}{\partial \beta}Y^TX\beta - \frac{\partial}{\partial \beta}\beta^TX^TY + \frac{\partial}{\partial \beta}\beta^TX^TX\beta = 0$$

Expanding to the derivatives of the individual terms

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial\beta}Y^T(Y - X\beta) - (X\beta)^T(Y - X\beta) = 0$$

Expanding to the derivatives of the individual terms

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial\beta}Y^T(Y - X\beta) - (X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

Expanding to the derivatives of the individual terms

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial\beta}Y^T(Y - X\beta) - (X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - Y^TX\beta - (X\beta)^T(Y - X\beta) = 0$$

Expanding to the derivatives of the individual terms

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial \beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial \beta}Y^T(Y - X\beta) - (X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - Y^TX\beta - (X\beta)^T(Y - X\beta) = 0$$

Expanding the second parenthesis
$(AB)^T = B^TA^T$

Expanding to the derivatives of the individual terms

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - \frac{\partial}{\partial \beta}Y^TX\beta - \frac{\partial}{\partial \beta}\beta^TX^TY + \frac{\partial}{\partial \beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial \beta}(Y - X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial \beta}Y^T(Y - X\beta) - (X\beta)^T(Y - X\beta) = 0$$

Expanding the first parenthesis

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - Y^TX\beta - (X\beta)^T(Y - X\beta) = 0$$

Expanding the second parenthesis
$(AB)^T = B^TA^T$

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta = 0$$

Expanding to the derivatives of the individual terms

$$\Rightarrow \frac{\partial}{\partial \beta}Y^TY - \frac{\partial}{\partial \beta}Y^TX\beta - \frac{\partial}{\partial \beta}\beta^TX^TY + \frac{\partial}{\partial \beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\frac{\partial}{\partial\beta}A = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\frac{\partial}{\partial\beta}A = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\frac{\partial}{\partial\beta}A\beta = A$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\frac{\partial}{\partial\beta}A\beta = A$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

Proof: $\beta^T X^T Y = (X^T Y)\beta$

$$\Rightarrow \frac{\partial}{\partial\beta}Y^T Y - \frac{\partial}{\partial\beta}Y^T X\beta - \frac{\partial}{\partial\beta}\beta^T X^T Y + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^T X\beta - \frac{\partial}{\partial\beta}\beta^T X^T Y + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial\beta}\beta^T X^T Y + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial\beta}(X^T Y)^T \beta + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

Proof: $\beta^T X^T Y = (X^T Y)\beta$

$$\beta^T X^T Y = (\beta^T X^T) Y$$
$$= (X\beta)^T Y$$

$$A^T B^T = (BA)^T$$

$$\Rightarrow \frac{\partial}{\partial \beta} Y^T Y - \frac{\partial}{\partial \beta} Y^T X\beta - \frac{\partial}{\partial \beta} \beta^T X^T Y + \frac{\partial}{\partial \beta} \beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial \beta} Y^T X\beta - \frac{\partial}{\partial \beta} \beta^T X^T Y + \frac{\partial}{\partial \beta} \beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} \beta^T X^T Y + \frac{\partial}{\partial \beta} \beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} (X^T Y)^T \beta + \frac{\partial}{\partial \beta} \beta^T X^T X\beta = 0$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1) \times (k+1)$ matrix
- β is a $(k+1) \times 1$ column vector

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y = (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric

Scalars are always symmetric

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric

  Scalars are always symmetric
- $\Rightarrow \beta^TX^TY$ is symmetric

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric
  Scalars are always symmetric
- $\Rightarrow \beta^TX^TY$ is symmetric
- if A is symmetric then $A = A^T$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric
  Scalars are always symmetric
- $\Rightarrow \beta^TX^TY$ is symmetric
- if A is symmetric then $A = A^T$
- therefore…

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric
  Scalars are always symmetric
- $\Rightarrow \beta^TX^TY$ is symmetric
- if A is symmetric then $A = A^T$
- therefore…

$$\beta^TX^TY = (\beta^TX^TY)^T$$
$$= (X^TY)^T\beta$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

Proof: $\beta^TX^TY = (X^TY)\beta$

$$\beta^TX^TY = (\beta^TX^T)Y$$
$$= (X\beta)^TY$$

- X is a $(n+1)\times(k+1)$ matrix
- β is a $(k+1)\times 1$ column vector
- $\Rightarrow X\beta$ is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^T$ is $1\times(n+1)$ row vector
- Y is a $(n+1)\times 1$ column vector
- $\Rightarrow (X\beta)^TY$ is a scalar (1×1)
- $\Rightarrow (X\beta)^TY$ is symmetric
  Scalars are always symmetric
- $\Rightarrow \beta^TX^TY$ is symmetric
- if A is symmetric then $A = A^T$
- therefore…

$$\beta^TX^TY = (\beta^TX^TY)^T$$
$$= (X^TY)^T\beta$$

Q.E.D

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$(AB)^T = B^TA^T$$

$$(A^T)^T = A$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow \frac{\partial}{\partial\beta}Y^TY - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - \frac{\partial}{\partial\beta}Y^TX\beta - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}\beta^TX^TY + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}(X^TY)^T\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$\Rightarrow 0 - Y^TX - \frac{\partial}{\partial\beta}Y^TX\beta + \frac{\partial}{\partial\beta}\beta^TX^TX\beta = 0$$

$$(AB)^T = B^TA^T$$

$$(A^T)^T = A$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} Y^T X \beta + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} Y^T X \beta + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

$$\frac{\partial}{\partial \beta} A\beta = A$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial\beta} Y^T X\beta + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\frac{\partial}{\partial\beta} A\beta = A$$

$$\Rightarrow 0 - Y^T X - Y^T X + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial\beta} Y^T X\beta + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\frac{\partial}{\partial\beta} A\beta = A$$

$$\Rightarrow 0 - Y^T X - Y^T X + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - Y^T X + 2\beta^T (X^T X) = 0$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} Y^T X \beta + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

$$\Rightarrow 0 - Y^T X - Y^T X + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

$$\Rightarrow 0 - Y^T X - Y^T X + 2\beta^T (X^T X) = 0$$

$$\frac{\partial}{\partial \beta} A\beta = A$$

$$\frac{\partial}{\partial \beta} \beta^T A \beta = \beta^T (A^T + A)$$

if A is symmetric…

$$\frac{\partial}{\partial \beta} \beta^T A \beta = \beta^T (A + A)$$

$$\frac{\partial}{\partial \beta} \beta^T A \beta = 2\beta^T A$$

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial \beta} Y^T X \beta + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

$$\frac{\partial}{\partial \beta} A\beta = A$$

$$\Rightarrow 0 - Y^T X - Y^T X + \frac{\partial}{\partial \beta} \beta^T X^T X \beta = 0$$

$$\Rightarrow 0 - Y^T X - Y^T X + 2\beta^T (X^T X) = 0$$

$$\Rightarrow -2Y^T X + 2\beta^T (X^T X) = 0$$

Combining the two terms

Multiple Regression: Compute the matrix β such that the SSR is minimized.

$$\Rightarrow 0 - Y^T X - \frac{\partial}{\partial\beta} Y^T X\beta + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\frac{\partial}{\partial\beta} A\beta = A$$

$$\Rightarrow 0 - Y^T X - Y^T X + \frac{\partial}{\partial\beta}\beta^T X^T X\beta = 0$$

$$\Rightarrow 0 - Y^T X - Y^T X + 2\beta^T (X^T X) = 0$$

$$\Rightarrow -2Y^T X + 2\beta^T (X^T X) = 0$$

Combining the two terms

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Adding $2Y^T X$ to both sides

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$\Rightarrow 2\beta^T X^T X = 2Y^T X$ Divide both sides by 2

$\Rightarrow \beta^T X^T X = Y^T X$ Transpose both sides

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Transpose both sides

$$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Transpose both sides

$$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$$

$$(AB)^T = B^T A^T$$

$$(A^T)^T = A$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Transpose both sides

$$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$$

$$(AB)^T = B^T A^T$$

$$(A^T)^T = A$$

$$\Rightarrow X^T X\beta = X^T Y$$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Transpose both sides

$$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$$

$$(AB)^T = B^T A^T$$

$$(A^T)^T = A$$

$$\Rightarrow X^T X\beta = X^T Y$$

Multiply both sides by $(X^T X)^{-1}$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$$\Rightarrow 2\beta^T X^T X = 2Y^T X$$

Divide both sides by 2

$$\Rightarrow \beta^T X^T X = Y^T X$$

Transpose both sides

$$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$$

$$(AB)^T = B^T A^T$$

$$(A^T)^T = A$$

$$\Rightarrow X^T X\beta = X^T Y$$

$$\Rightarrow \beta = (X^T X)^{-1}(X^T Y)$$

Multiply both sides by $(X^T X)^{-1}$

Multiple Regression: Compute the matrix β such that the MSE is minimized.

$\Rightarrow 2\beta^T X^T X = 2Y^T X$

Divide both sides by 2

$\Rightarrow \beta^T X^T X = Y^T X$

Transpose both sides

$\Rightarrow (\beta^T X^T X)^T = (Y^T X)^T$

$(AB)^T = B^T A^T$

$(A^T)^T = A$

$\Rightarrow X^T X\beta = X^T Y$

$\Rightarrow \beta = (X^T X)^{-1}(X^T Y)$

Multiply both sides by $(X^T X)^{-1}$

Q.E.D

Related Tutorials & Textbooks

Simple Linear Regression

A statistical technique of making predictions from data. The tutorial introduces a linear model in two dimensions and uses that model to predict the value of one dependent variable given one independent variable.

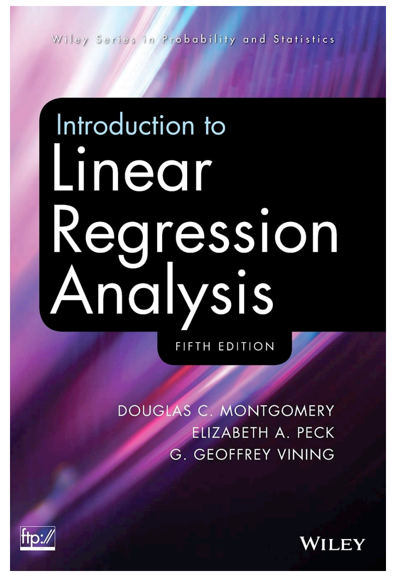
Multiple Regression

Multiple regression extends the two dimensional linear model introduced in Simple Linear Regression to k + 1 dimensions with one dependent variable, k independent variables and k+1 parameters.

Gradient Descent for Multiple Regression

Gradient Descent algorithm for multiple regression and how it can be used to optimize k + 1 parameters for a Linear model in multiple dimensions.

Recommended Textbooks



Introduction to Linear Regression Analysis

by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining

For a complete list of tutorials see:
https://arrsingh.com/ai-tutorials