UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

> Reinventa el mundo <

# LINEAR REGRESSION

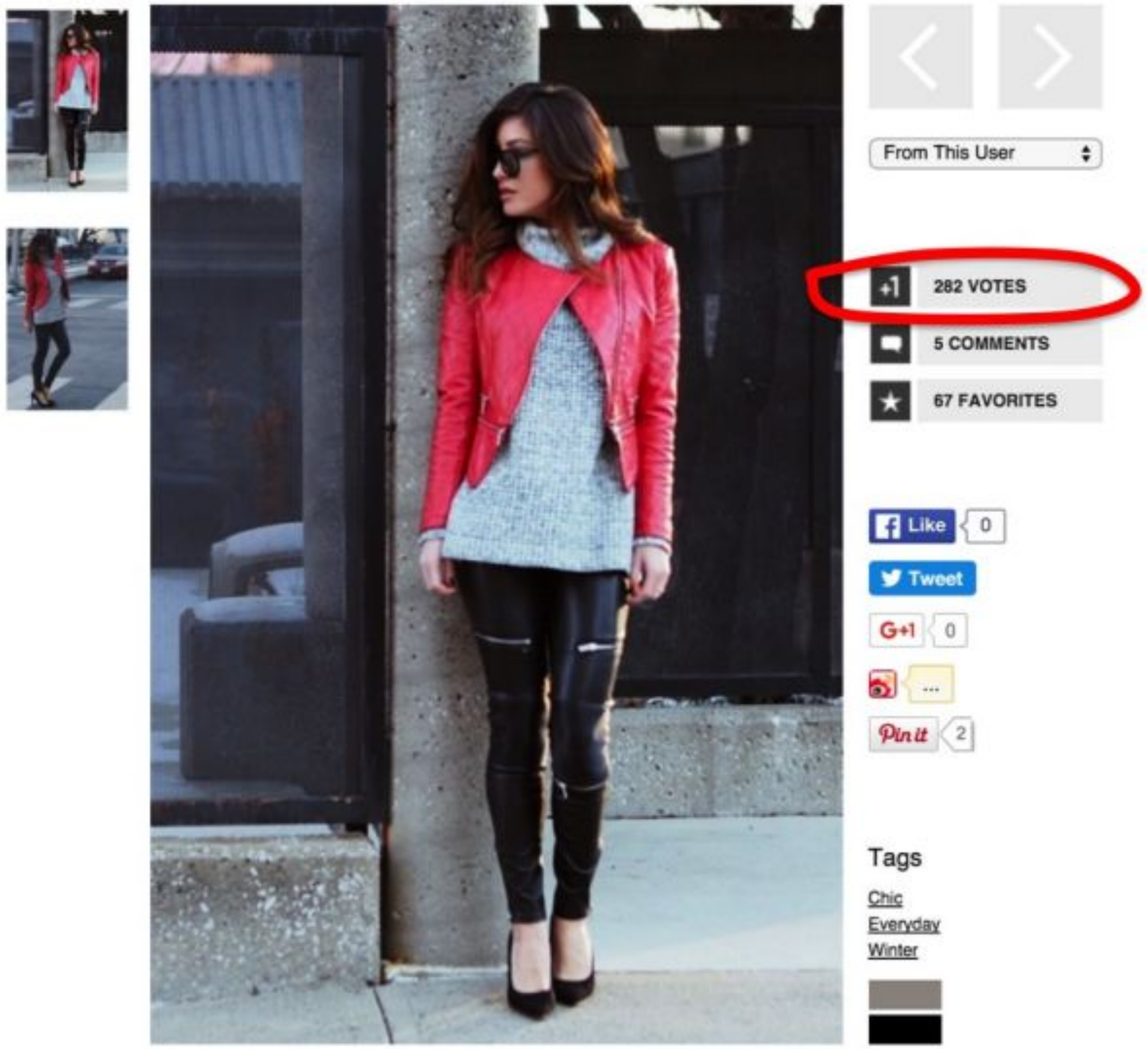
*Ariana Villegas*

TRANSFORMATEC

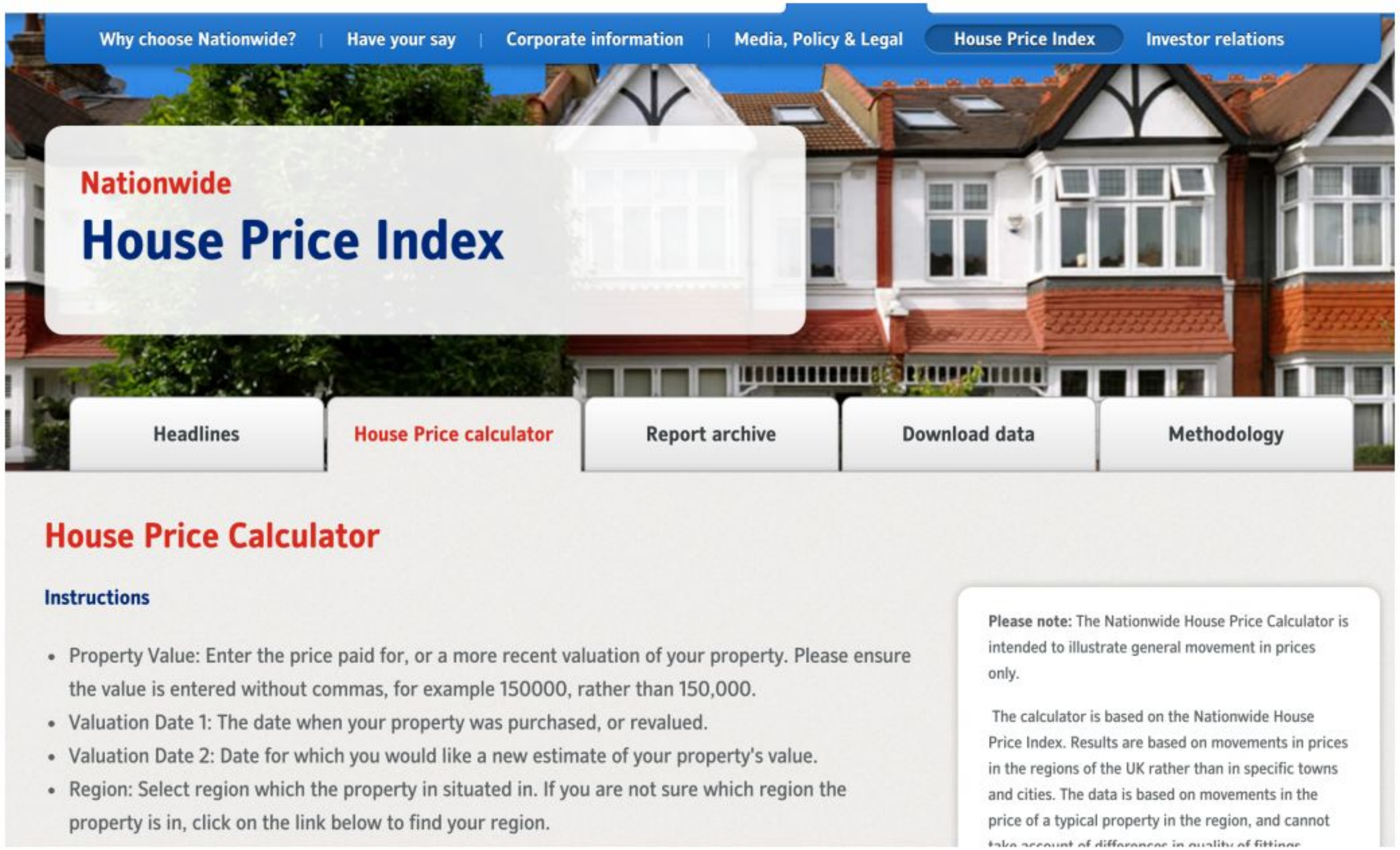
# What should I watch?


Can we predict this?

# How many followers will I get?



# Predict the price of the house



# What do all these problems have in common?

UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

# 1.

## Regression

*Examples & Statistics*


TRANSFORMATEC

# REGRESSION aka Fitting Curves to Data

**Classification:** given point x, predict class (often binary)

**Regression:** given point x, predict a numerical value

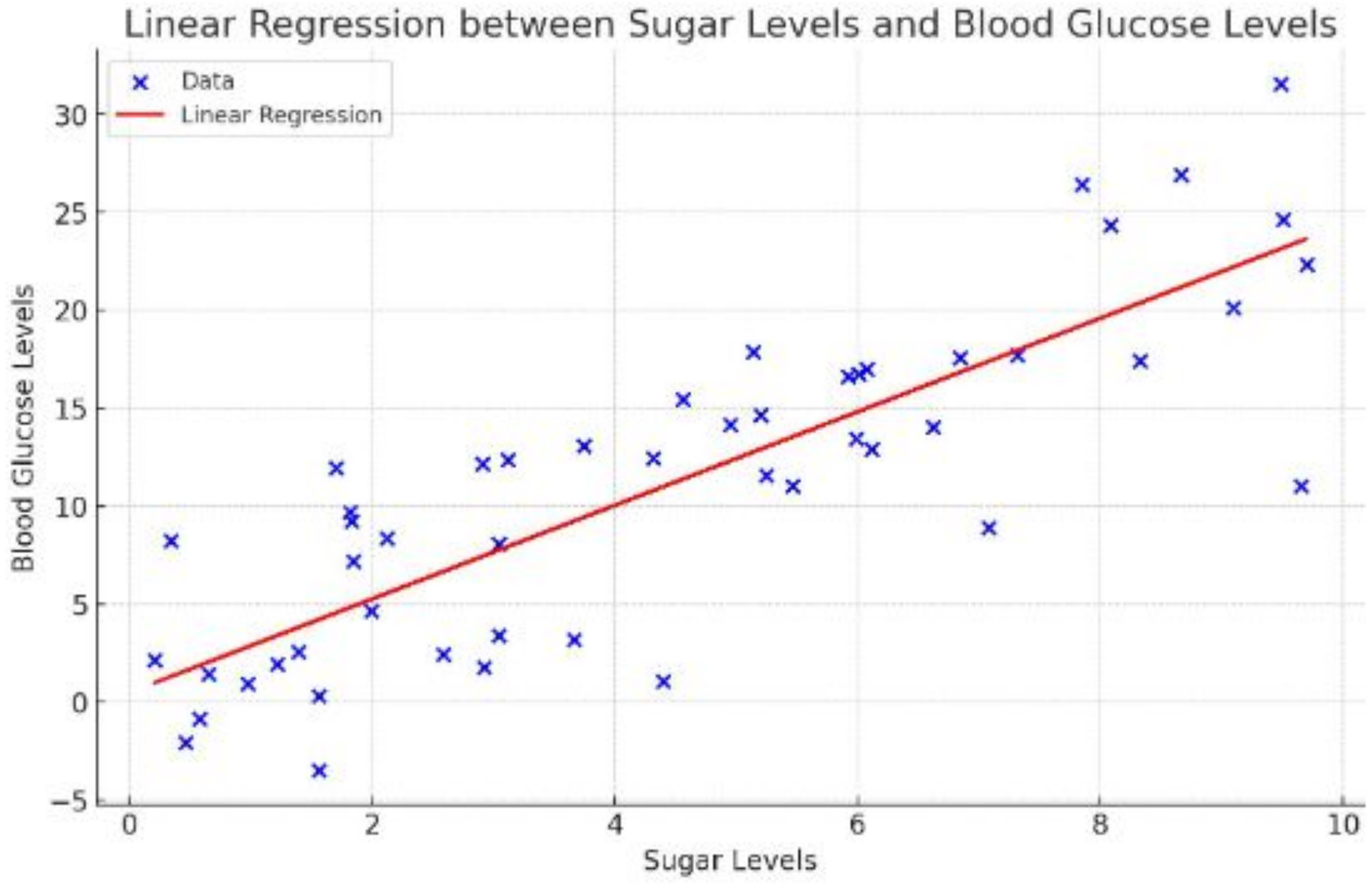
What will be the temperature tomorrow?


Fahrenheit

# REGRESSION aka Fitting Curves to Data

**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?

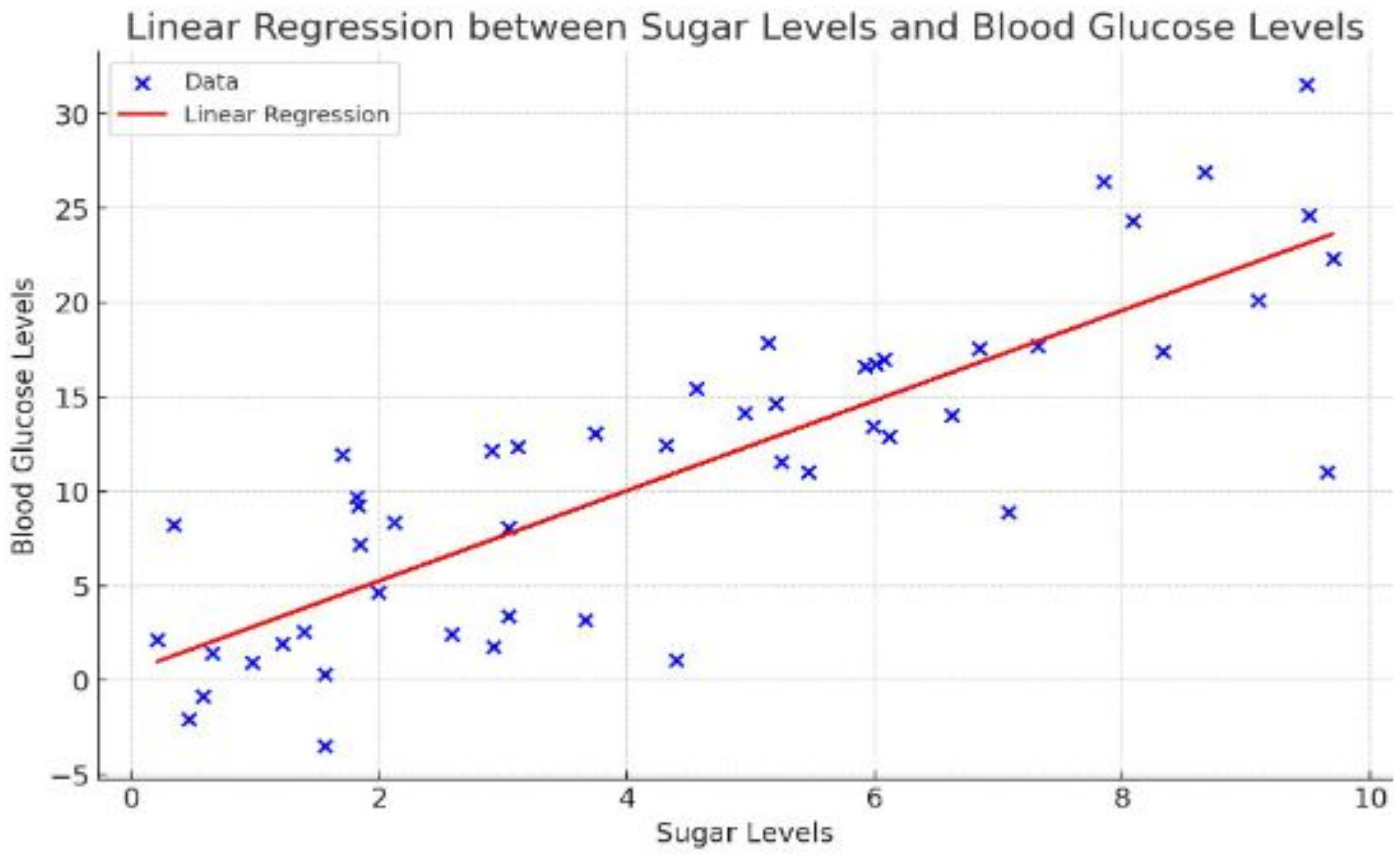



| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

# REGRESSION aka Fitting Curves to Data

**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?




**Input (x)** **Output (y)**

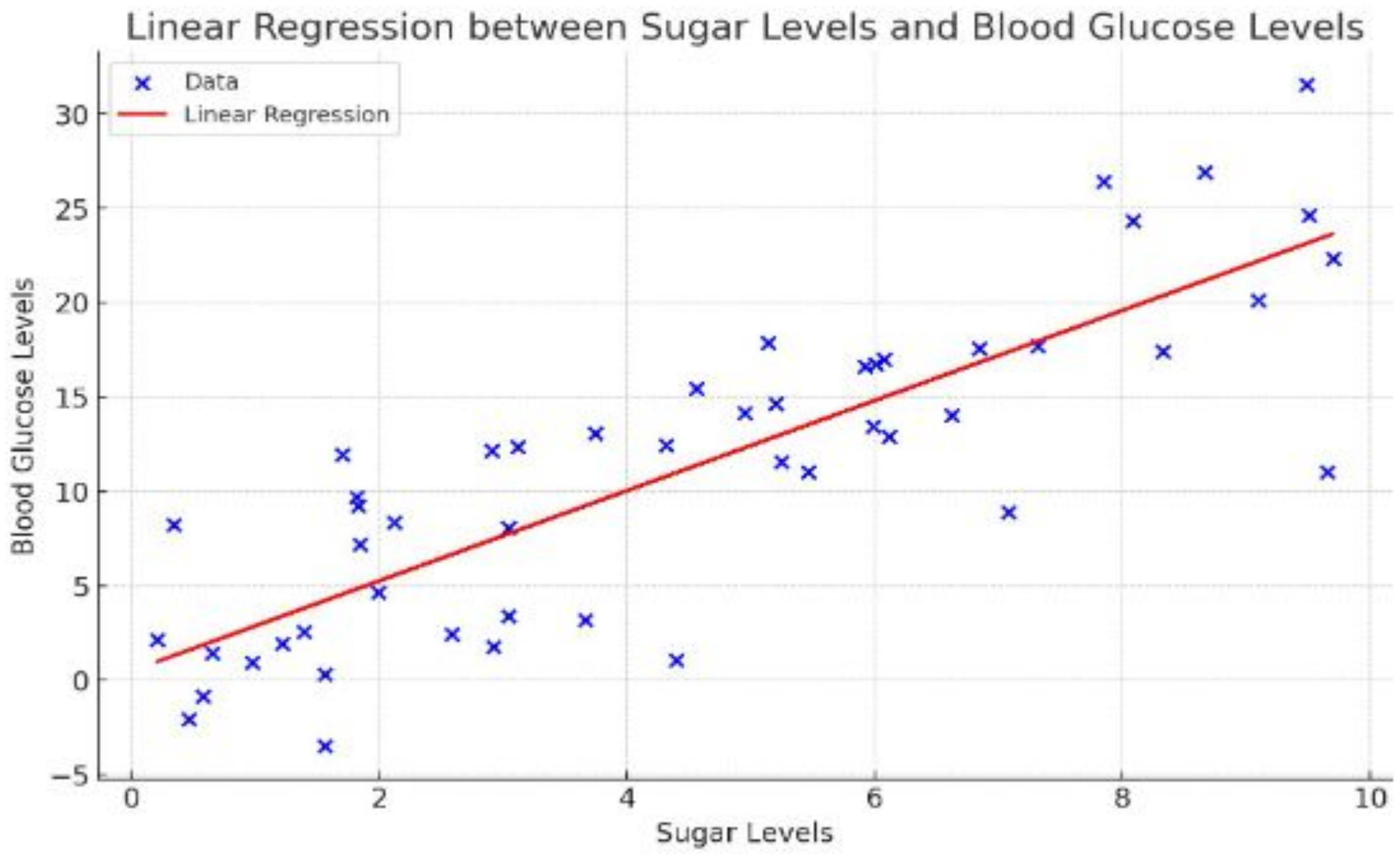
| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

$$f(x) = mx + b$$

# REGRESSION aka Fitting Curves to Data

**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?




Input (x) Output (y)

| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

$f(x) = mx + b$

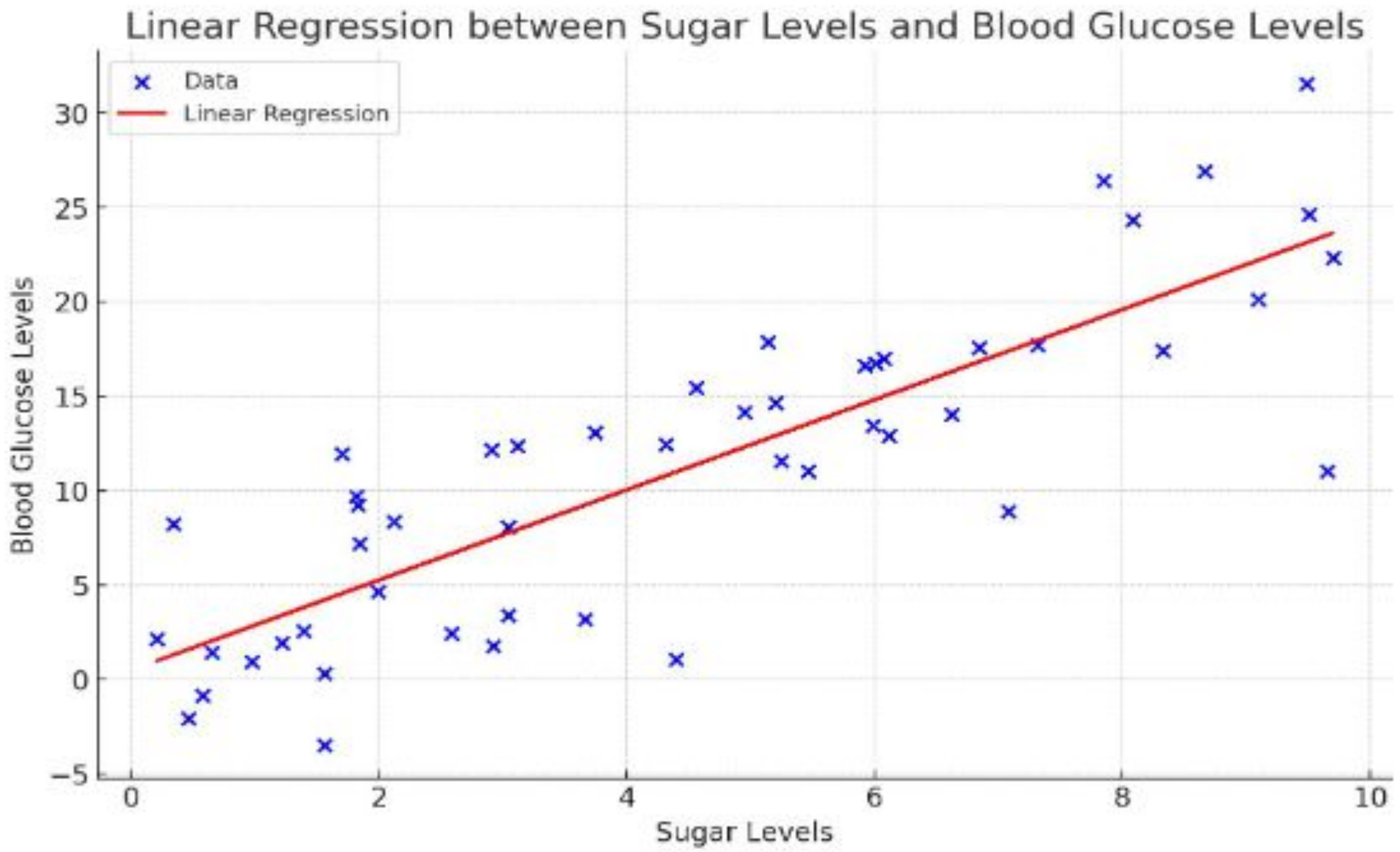
**Model**

# REGRESSION aka Fitting Curves to Data

**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?

**Error?**




**Input (x)** **Output (y)**

| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

$f(x) = mx + b$

**Model**

# REGRESSION aka Fitting Curves to Data

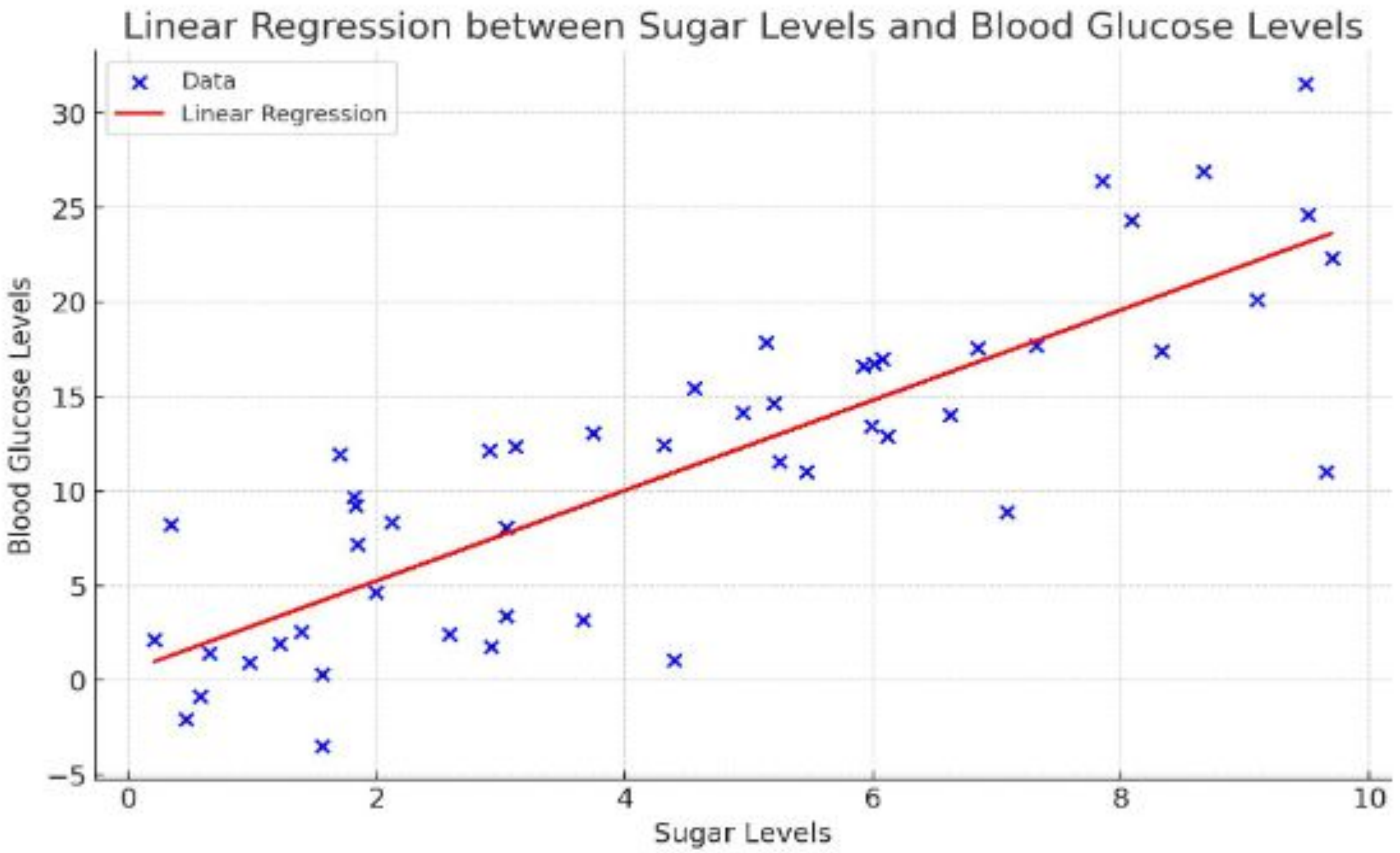
**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?

**Error**

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$




**Input (x)** **Output (y)**

| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

f(x) = mx + b

**Model**

# REGRESSION aka Fitting Curves to Data

What do I need in order to predict **these outputs**?

**Least Squares Regression:** y = mx + b

$$m = \frac{N\,\Sigma(xy) - \Sigma x\,\Sigma y}{N\,\Sigma(x^2) - (\Sigma x)^2} \qquad b = \frac{\Sigma y - m\,\Sigma x}{N}$$

$$\mathbf{y} = mx + b$$

# Loss Function vs Metrics

*Simply put, loss function is for machines, and metric is for humans.*

Loss function is what the machine tries to minimize in order to optimize the machine learning model.

Metrics are utilized by people to evaluate the performance of machine learning models and has nothing to do with the optimization process.

**Can they be used interchangeably?**

A **loss function** can be used as a metric, but the opposite isn't always true.

This is due to an important characteristic of loss functions: they must be differentiable.

# Condition of a Metric

**Definition 7.1.** A metric $d$ on a set $X$ is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y \in X$:

(1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;

(2) $d(x, y) = d(y, x)$ (symmetry);

(3) $d(x, y) \leq d(x, z) + d(z, x)$ (triangle inequality).

A metric space $(X, d)$ is a set $X$ with a metric $d$ defined on $X$.

*Metric spaces by UCDavis*

# Condition of a Metric

**Definition 7.1.** A metric $d$ on a set $X$ is a function $d : X \times X \to \mathbb{R}$ such that for all $x, y \in X$:

(1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
(2) $d(x, y) = d(y, x)$ (symmetry);
(3) $d(x, y) \leq d(x, z) + d(z, x)$ (triangle inequality).

A metric space $(X, d)$ is a set $X$ with a metric $d$ defined on $X$.

*Metric spaces by UCDavis*

# Metrics for Regression

$$\text{MAPE} = \frac{\Sigma \frac{|A - F|}{A} \times 100}{N}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)}$$

$$\textbf{R2 Squared} = 1 - \frac{SSr}{SSm}$$

SSr = Squared sum error of regression line

SSm = Squared sum error of mean line

# 2.

## **Linear Regression**
*Machine Learning*

# REGRESSION aka Fitting Curves to Data

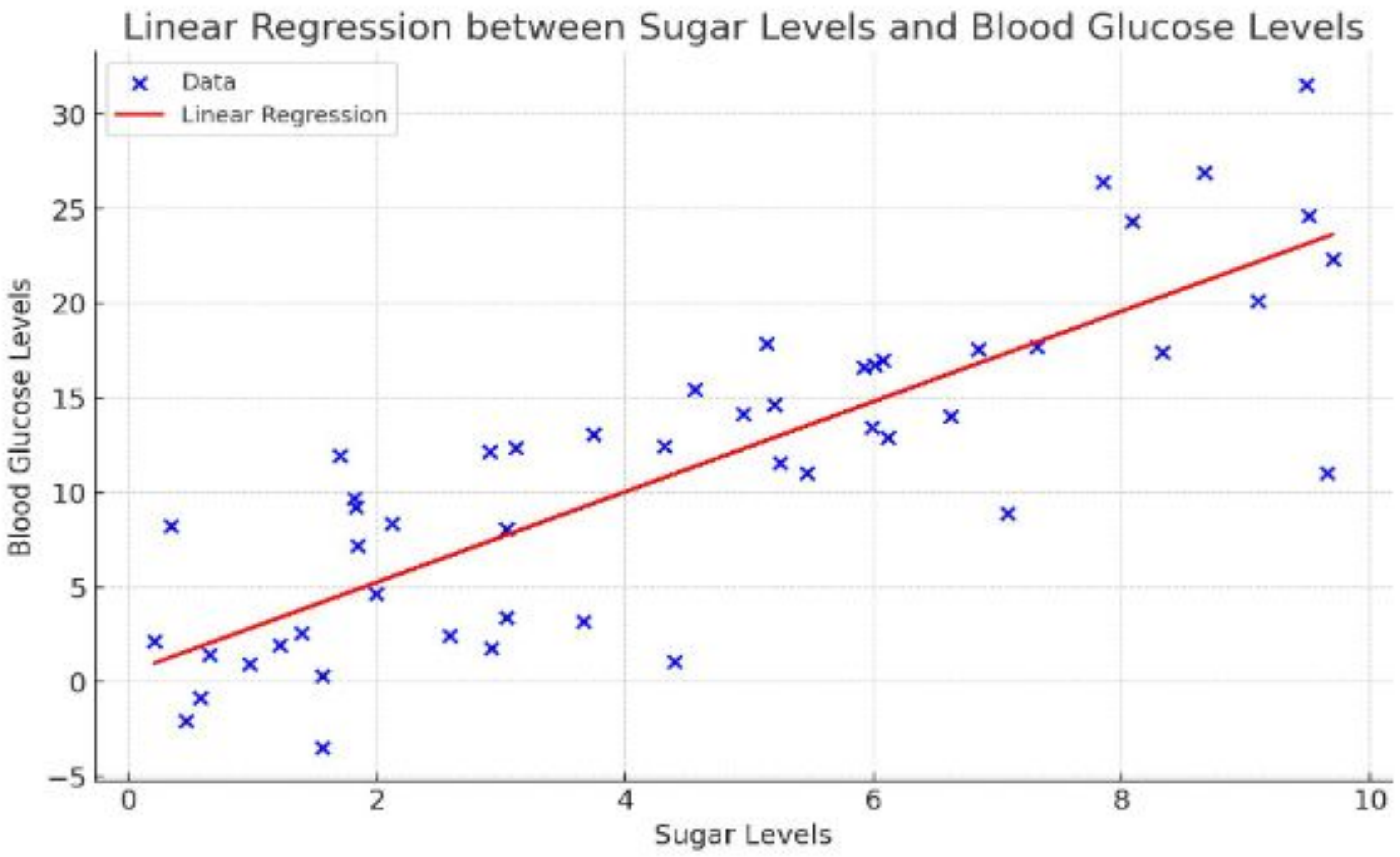
**Regression:** given point x, predict a **numerical value**

What do I need in order to predict **these outputs**?

**Error**

MSE?

MAE?



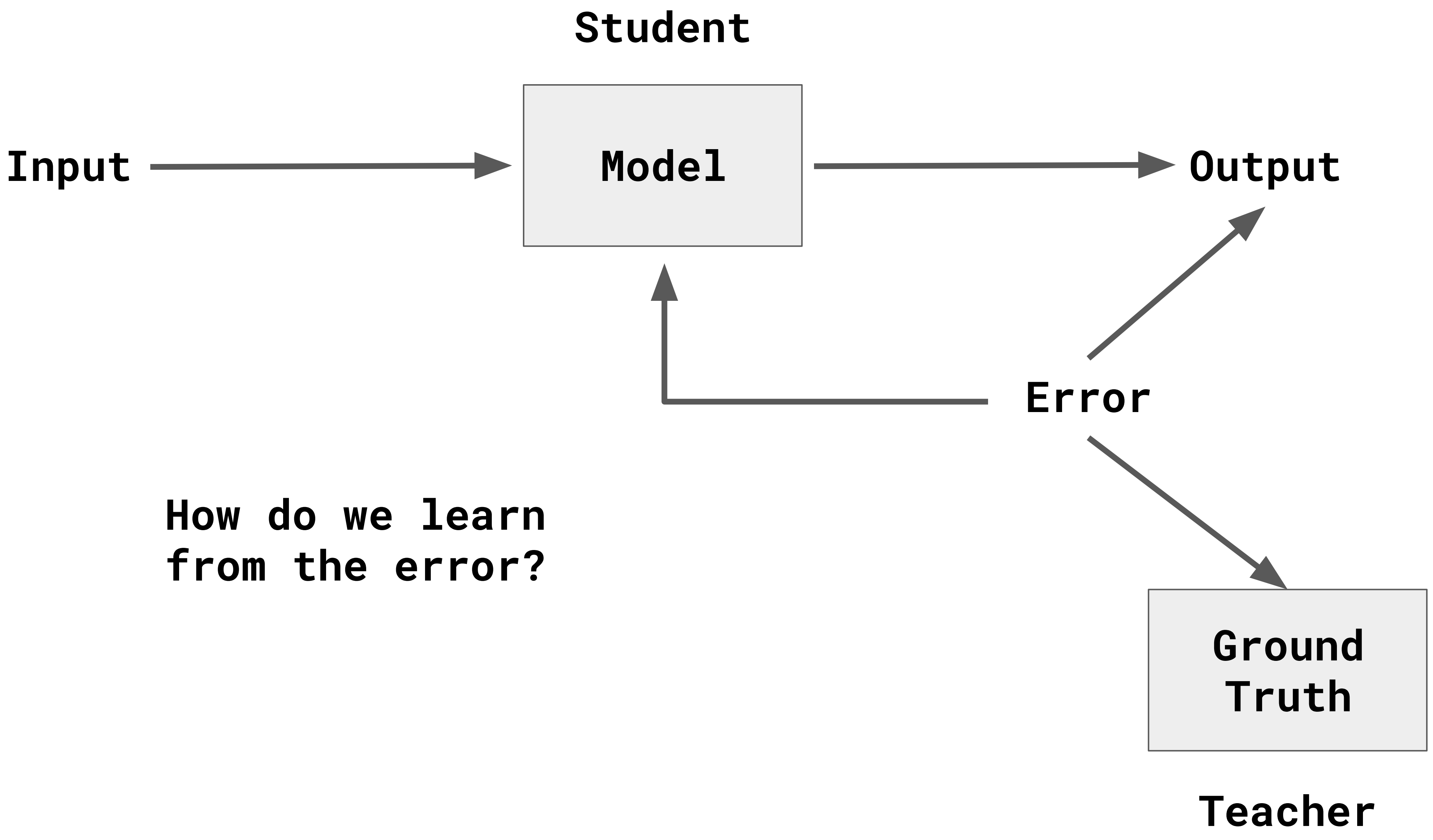

**Input (x)** **Output (y)**

| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

f(x) = ax + b => f(x) = wx + b

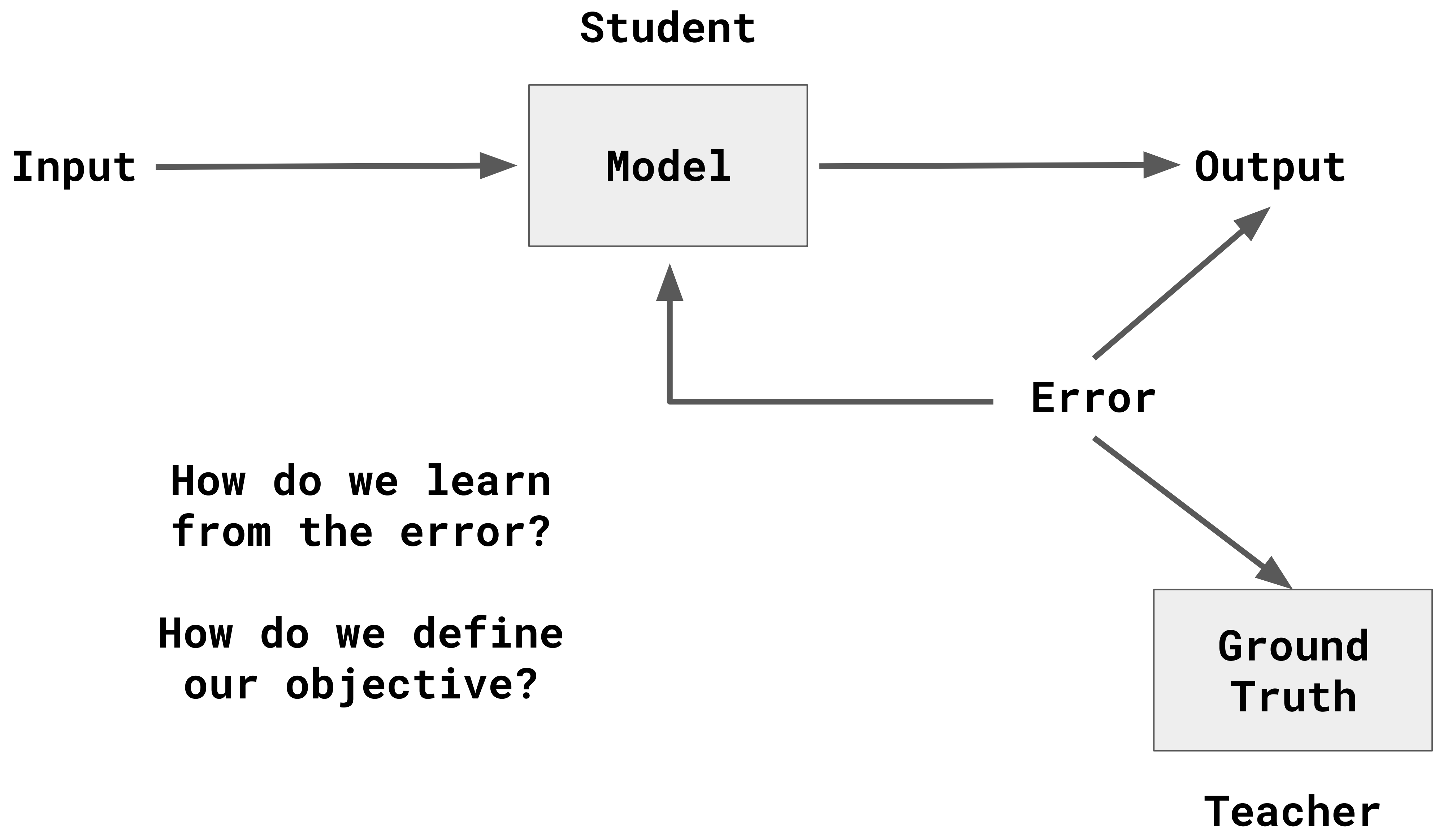
**Model**

# Regression




How do we learn from the error?

# Regression




How do we learn from the error?

How do we define our objective?

# Regression

Student

Input → Model → Output

Error

Ground Truth

Teacher

How do we learn from the error?

How do we define our objective?

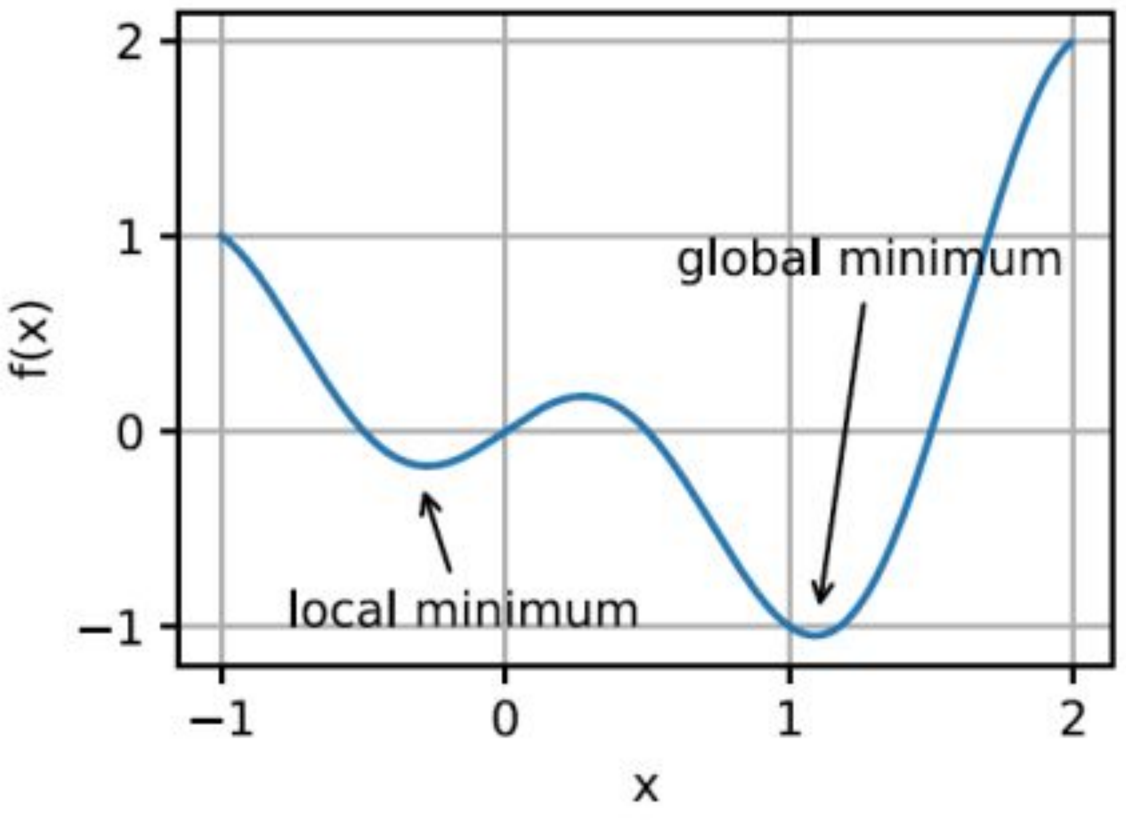
Minimize the error

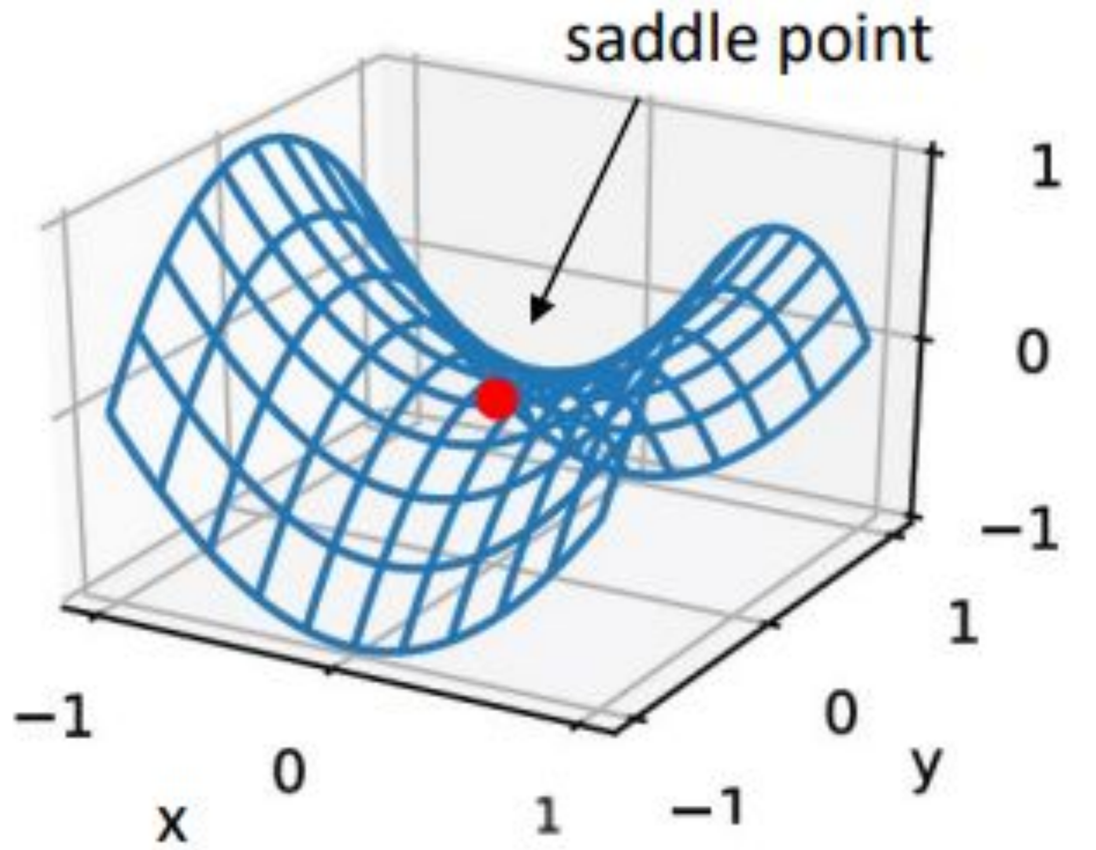
Gradient Descent

# Gradient Descent - Optimization

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$$

$$x_1 = x_1 - \alpha \frac{\partial f(x)}{\partial x_1} \quad x_2 = x_2 - \alpha \frac{\partial f(x)}{\partial x_2}$$

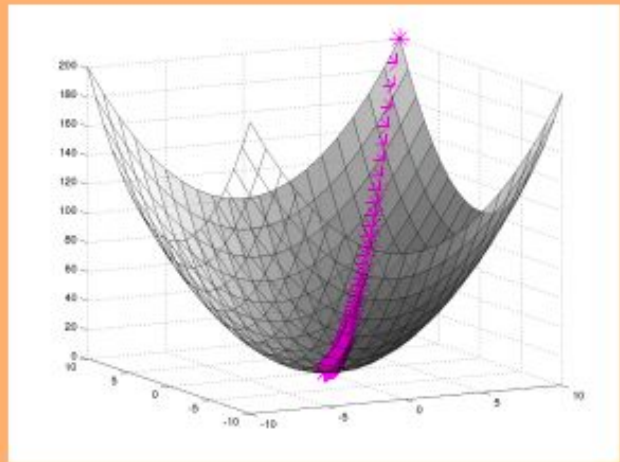


# Gradient Descent

**Algorithm 1** Gradient Descent

**procedure** GD($\mathcal{D}$, $\boldsymbol{\theta}^{(0)}$)
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^{(0)}$
    **while** not converged **do**
        $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \gamma \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$
    **return** $\boldsymbol{\theta}$



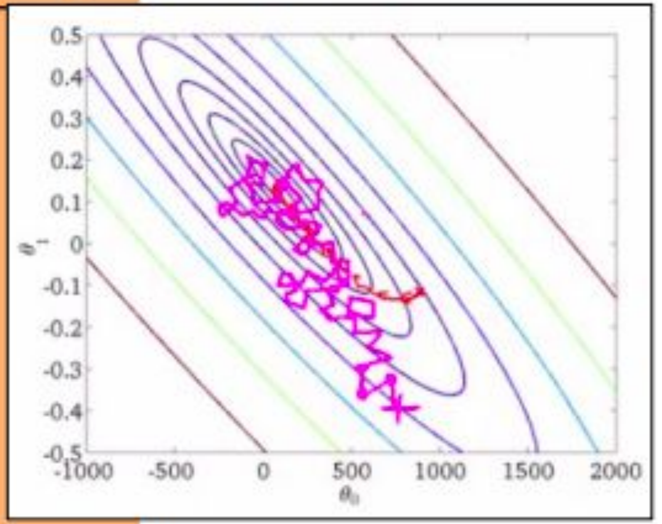
$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$$

$$x_1 = x_1 - \alpha \frac{\partial f(x)}{\partial x_1} \quad x_2 = x_2 - \alpha \frac{\partial f(x)}{\partial x_2}$$

# Stochastic Gradient Descent

**Algorithm 1** Stochastic Gradient Descent (SGD)

**procedure** SGD($\mathcal{D}, \boldsymbol{\theta}^{(0)}$)
    $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^{(0)}$
    **while** not converged **do**
        **for** $i \in \text{shuffle}(\{1, 2, \ldots, N\})$ **do**
            $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \gamma \nabla_{\boldsymbol{\theta}} J^{(i)}(\boldsymbol{\theta})$
    **return** $\boldsymbol{\theta}$




$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$$

$$x_1 = x_1 - \alpha \frac{\partial f(x)}{\partial x_1} \qquad x_2 = x_2 - \alpha \frac{\partial f(x)}{\partial x_2}$$

# Linear Regression

Model Prediction / Output → f(x) = y' = wx + b

y ← Ground Truth / Label

Minimize the error (MSE, MAE)

Loss = MSE = 1/n ∑(f(x) - y)²

$$w = w - \alpha\frac{\partial Loss}{\partial w}$$

$$b = b - \alpha\frac{\partial Loss}{\partial b}$$

# Linear Regression

Minimize the error (MSE, MAE)

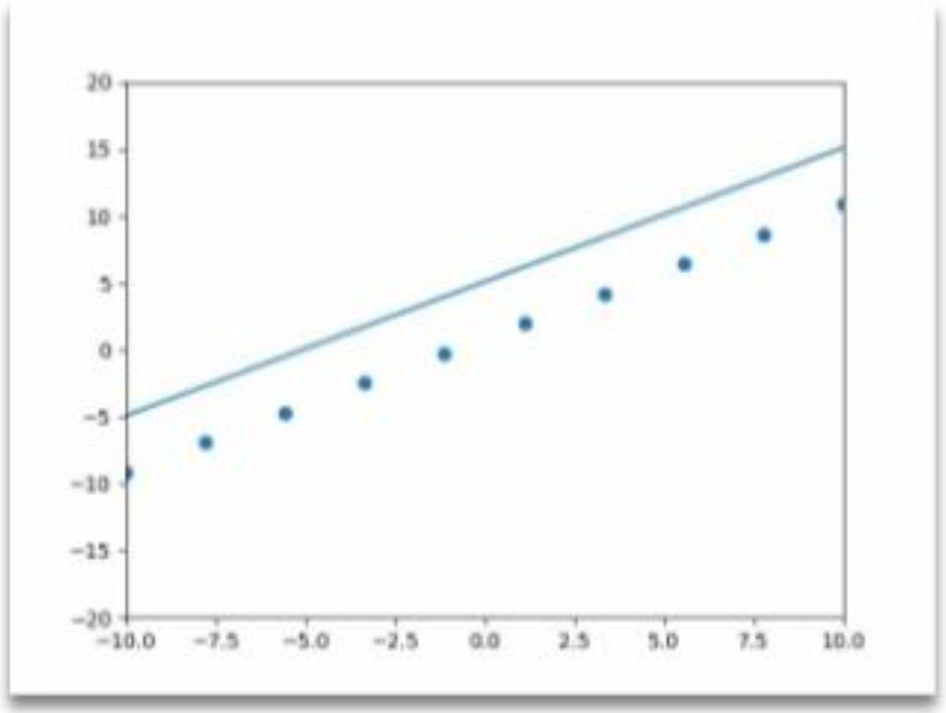
Loss = MSE = 1/2n ∑(f(x) - y)²

$$w = w - \alpha \frac{\partial Loss}{\partial w} = w - \alpha \frac{1}{n} \sum (f(x) - y)x$$

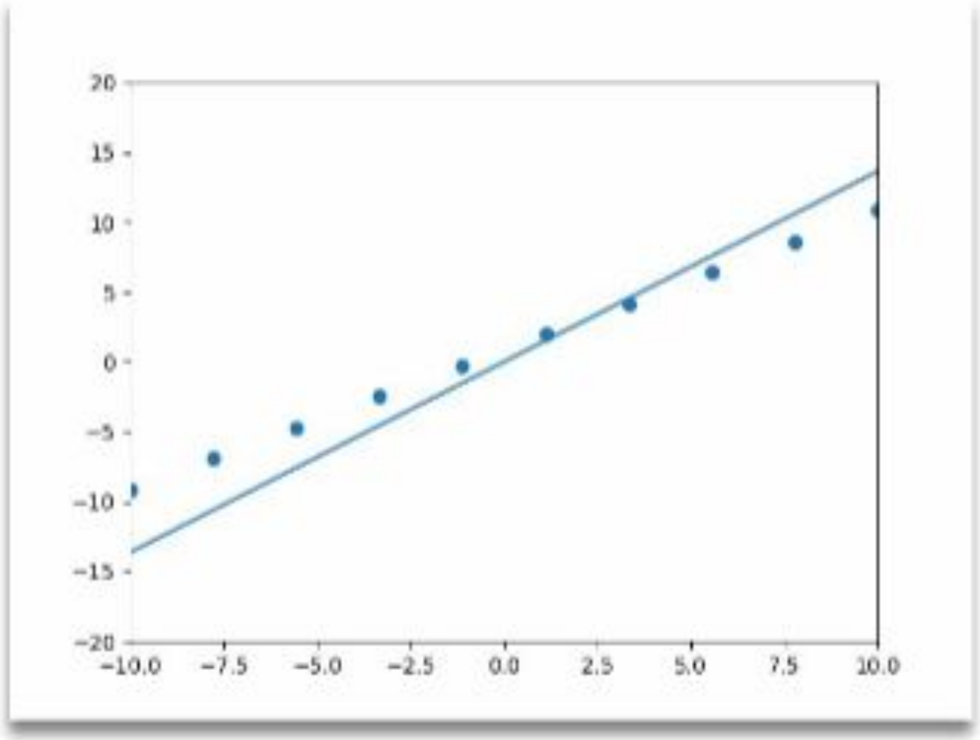
$$b = b - \alpha \frac{\partial Loss}{\partial b} = b - \alpha \frac{1}{n} \sum (f(x) - y)$$

# Linear Regression

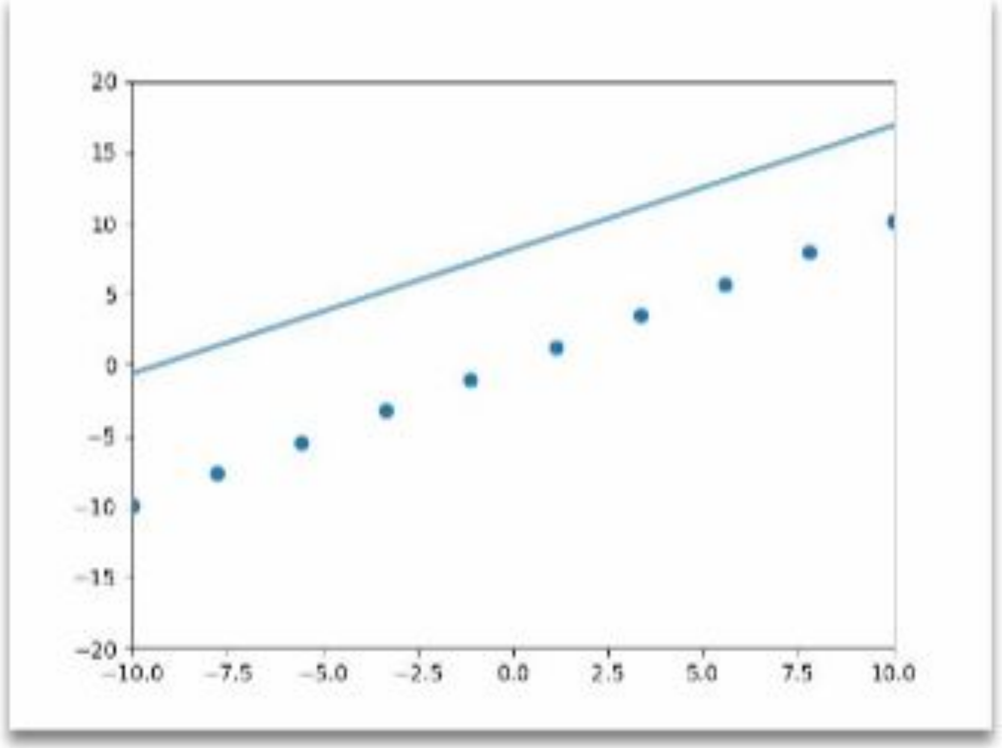
$$L = xw + \boxed{b}$$




$$L = x\boxed{w} + b$$




$$L = x\boxed{w} + \boxed{b}$$



# Multidimensional Linear Regression

Imagine now we want to predict the median house price from these **multi-dimensional** observations

Each house is a data point **n**, with observations indexed by **j**:

$$\mathbf{x}^{(n)} = \left(x_1^{(n)}, \cdots, x_j^{(n)}, \cdots, x_d^{(n)}\right)$$

We can incorporate the bias $\mathbf{w_0}$ into $\mathbf{w}$, by using $\mathbf{x_0 = 1}$, then

$$y(\mathbf{x}) = w_0 + \sum_{j=1}^{d} w_j x_j = \mathbf{w}^T \mathbf{x}$$

We can then solve for $w = (w_0, w_1, \cdot \cdot \cdot, w_d)$. How?

# Multidimensional Linear Regression

Imagine now we want to predict the median house price from these **multi-dimensional** observations

Each house is a data point **n**, with observations indexed by **j**:

$$\mathbf{x}^{(n)} = \left(x_1^{(n)}, \cdots, x_j^{(n)}, \cdots, x_d^{(n)}\right)$$

We can incorporate the bias $\mathbf{w_0}$ into $\mathbf{w}$, by using $\mathbf{x_0 = 1}$, then

$$y(\mathbf{x}) = w_0 + \sum_{j=1}^{d} w_j x_j = \mathbf{w}^T \mathbf{x}$$

We can then solve for $w = (w_0, w_1, \cdot \cdot \cdot, w_d)$. How?

We can use gradient descent to solve for each **coefficient**, or compute **w** analytically (how does the solution change?)

# Multidimensional Linear Regression

Imagine now we want to predict the median house price from these **multi-dimensional** observations

Each house is a data point **n**, with observations indexed by **j**:

$$\mathbf{x}^{(n)} = \left(x_1^{(n)}, \cdots, x_j^{(n)}, \cdots, x_d^{(n)}\right)$$

We can use gradient descent to solve for each **coefficient**, or compute **w** analytically (how does the solution change?)

$$\vec{\beta} = (X^T X)^{-1} X^T \vec{y}$$

$$\begin{bmatrix}\beta_0 \\ \beta_1\end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ \sum X_i y_i \end{bmatrix}$$

UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

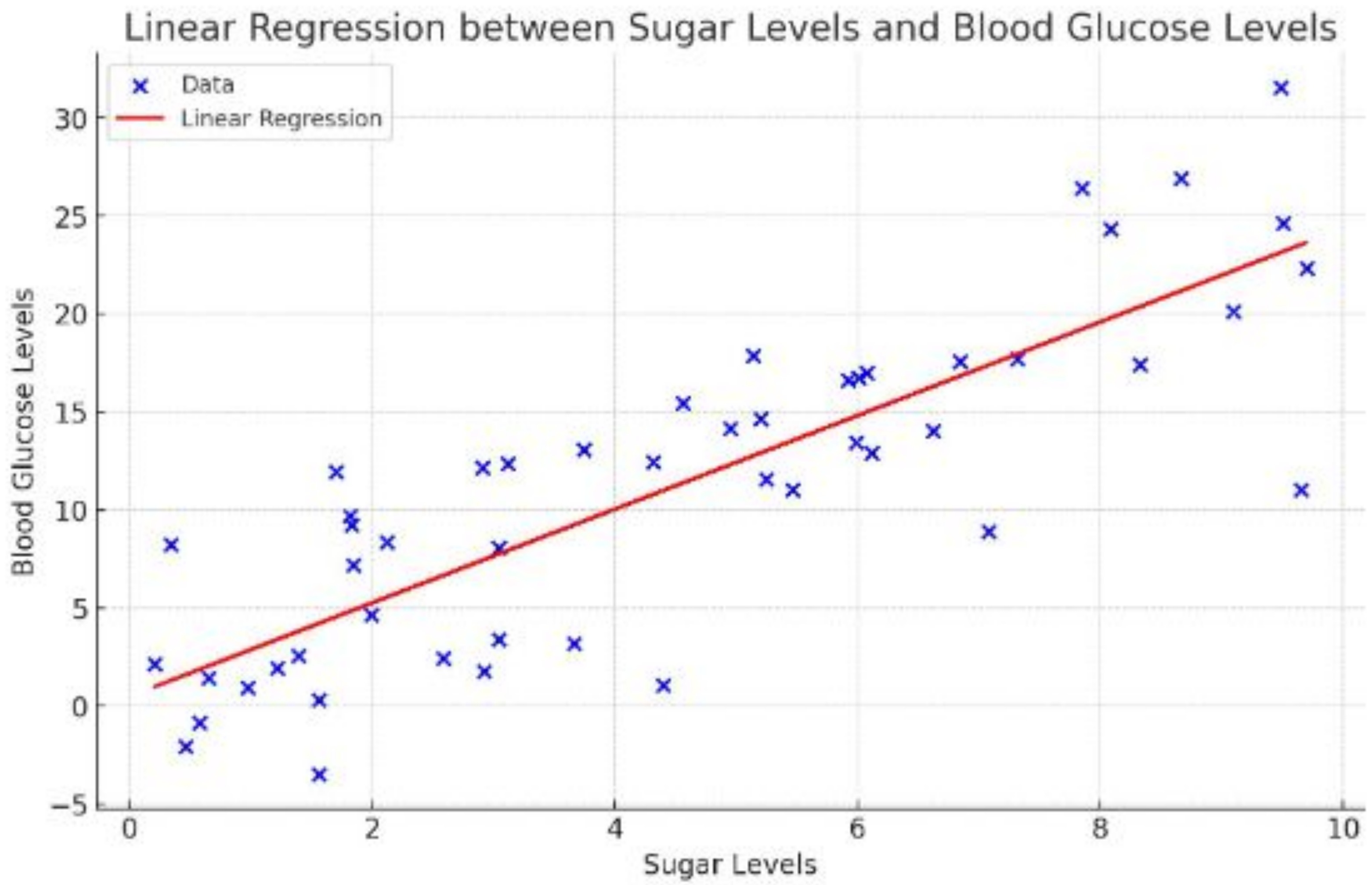
# 3.

## Linear Regression
*Example*


TRANSFORMATEC

https://www.mladdict.com/linear-regression-simulator

# Home Exercise




| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

f(x) = wx + b     w = 0     b = 0     a = 0.1

# Home Exercise

$f(x) = wx + b$ $w = 0$ $b = 0$ $a = 0.1$

| Sugar Levels | Blood Glucose Levels |
|---|---|
| 3.75 | 13.06 |
| 9.51 | 24.62 |
| 7.32 | 17.72 |
| 5.99 | 13.46 |
| 1.56 | -3.49 |

# More Exercises?

MACHINE LEARNING WORKOUT: Chapter 5


UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA
> Reinventa el mundo <
TRANSFORMATEC