

LINEAR REGRESSION

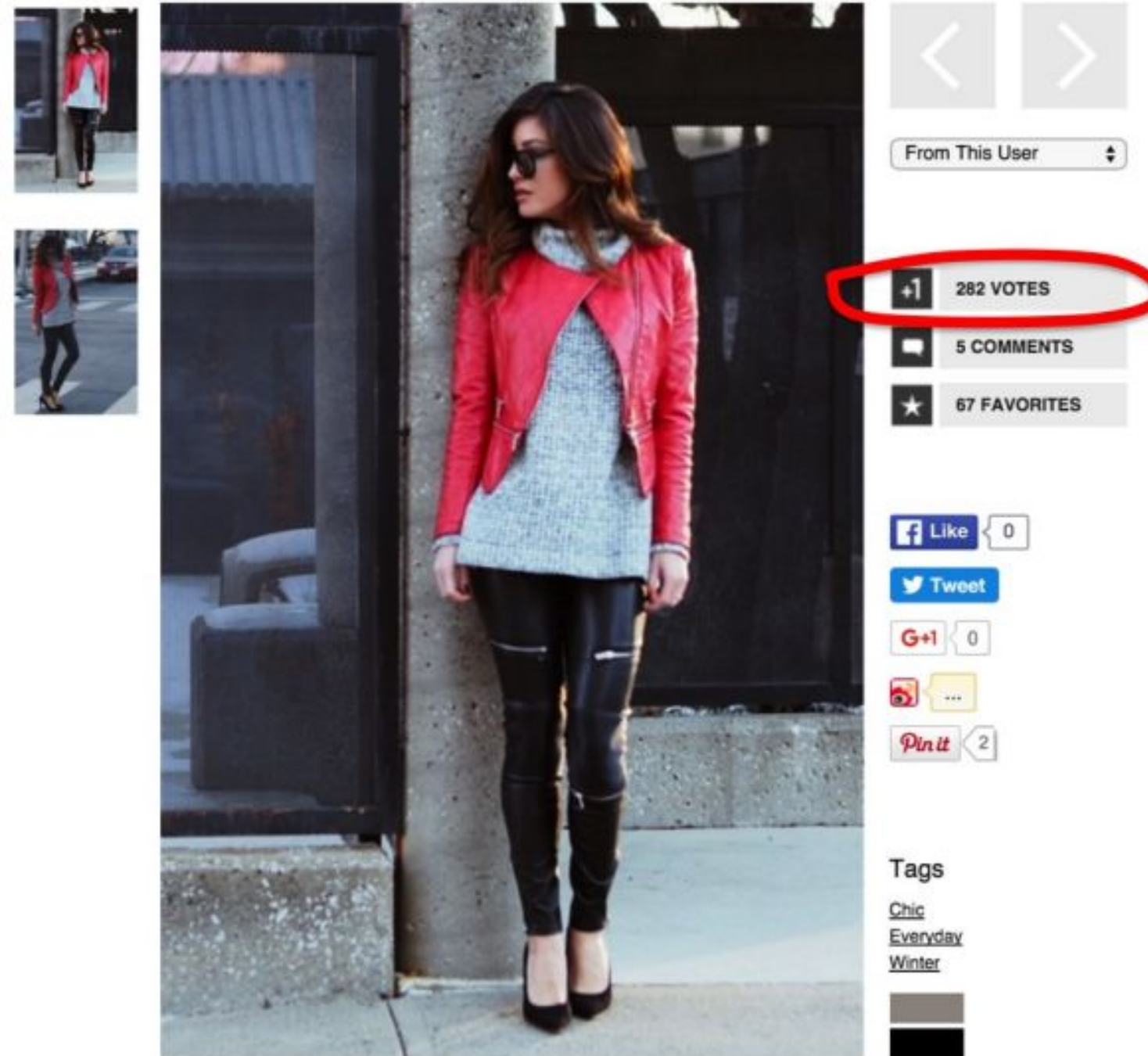
Ariana Villegas

What should I watch?

The screenshot shows the IMDb page for the movie 'The Martian' (2015). The movie poster on the left features Matt Damon in a space helmet with the text 'BRING HIM HOME' and 'THE MARTIAN'. To the right of the poster, the movie title 'The Martian (2015)' is displayed, followed by its rating 'PG-13', runtime '144 min', genres 'Adventure, Comedy, Drama', and release date '2 October 2015 (USA)'. A yellow star with the number '8.1' is circled in blue. To the right of the star, the text 'Your rating: ★★★★★★ -/10' is shown, followed by 'Ratings: 8.1/10 from 271,829 users' and 'Metascore: 80/100'. Below this, it says 'Reviews: 750 user | 499 critic | 46 from Metacritic.com'. The synopsis follows: 'During a manned mission to Mars, Astronaut Mark Watney is presumed dead after a fierce storm and left behind by his crew. But Watney has survived and finds himself stranded and alone on the hostile planet. With only meager supplies, he must draw upon his ingenuity, wit and spirit to subsist and find a way to signal to Earth that he is alive.' The director is listed as 'Ridley Scott' and the writers as 'Drew Goddard (screenplay), Andy Weir (book)'. The stars are 'Matt Damon, Jessica Chastain, Kristen Wiig'. At the bottom, there are buttons for '+ Watchlist', 'Watch Trailer', and 'Share...'. A blue arrow points from the circled star to the text 'Can we predict this?'.

Can we predict this?

How many followers will I get?



Predict the price of the house

The screenshot shows the Nationwide House Price Index website. At the top is a blue navigation bar with links: 'Why choose Nationwide?', 'Have your say', 'Corporate information', 'Media, Policy & Legal', 'House Price Index' (highlighted), and 'Investor relations'. Below the navigation bar is a large image of a row of houses. Overlaid on this image is a white box with the text 'Nationwide House Price Index'. Below the image is a row of five buttons: 'Headlines', 'House Price calculator' (highlighted), 'Report archive', 'Download data', and 'Methodology'. Below the buttons is the 'House Price Calculator' section. It has a sub-header 'Instructions' followed by a list of instructions: 'Property Value: Enter the price paid for, or a more recent valuation of your property. Please ensure the value is entered without commas, for example 150000, rather than 150,000.', 'Valuation Date 1: The date when your property was purchased, or revalued.', 'Valuation Date 2: Date for which you would like a new estimate of your property's value.', and 'Region: Select region which the property is situated in. If you are not sure which region the property is in, click on the link below to find your region.' To the right of the instructions is a 'Please note' box stating: 'The Nationwide House Price Calculator is intended to illustrate general movement in prices only. The calculator is based on the Nationwide House Price Index. Results are based on movements in prices in the regions of the UK rather than in specific towns and cities. The data is based on movements in the price of a typical property in the region, and cannot take account of differences in quality of fittings.'



**What do all these problems
have in common?**

1.



Regression

Examples & Statistics

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?

84°



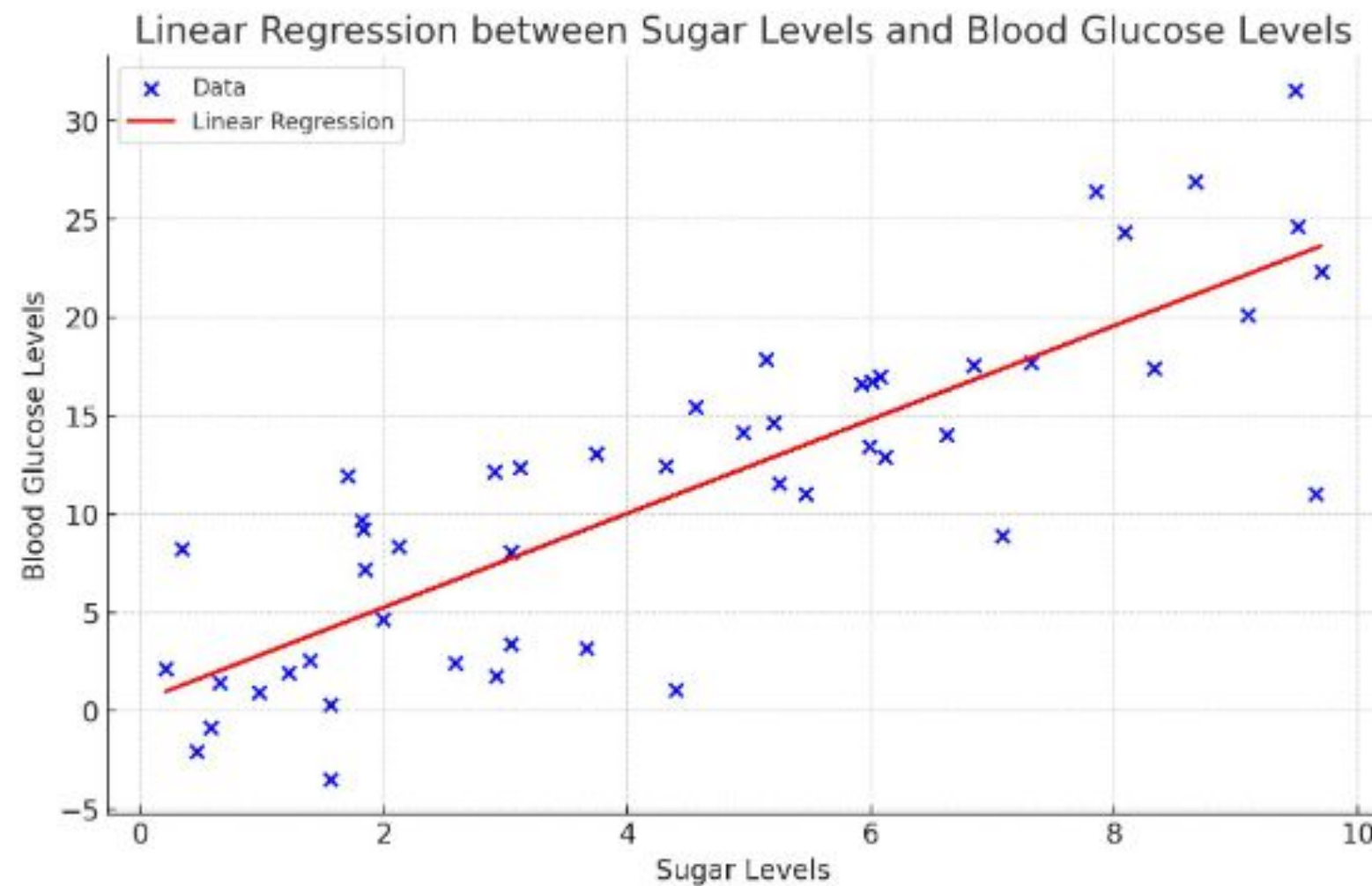
Fahrenheit



REGRESSION aka Fitting Curves to Data

Regression: given point x , predict a numerical value

What do I need in order to predict these outputs?

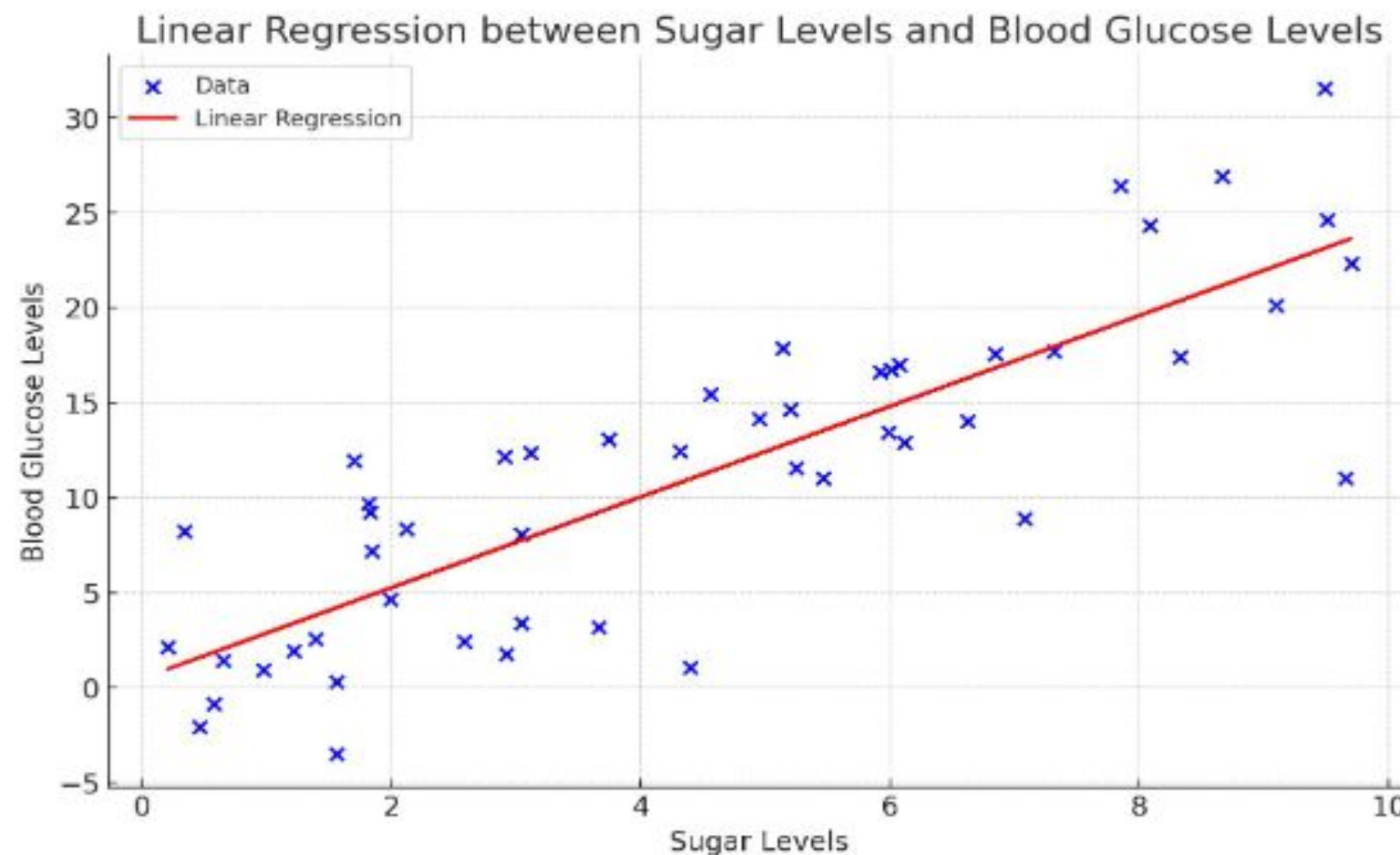


Sugar Levels	Blood Glucose Levels
3.75	13.06
9.51	24.62
7.32	17.72
5.99	13.46
1.56	-3.49

REGRESSION aka Fitting Curves to Data

Regression: given point x , predict a numerical value

What do I need in order to predict these outputs?



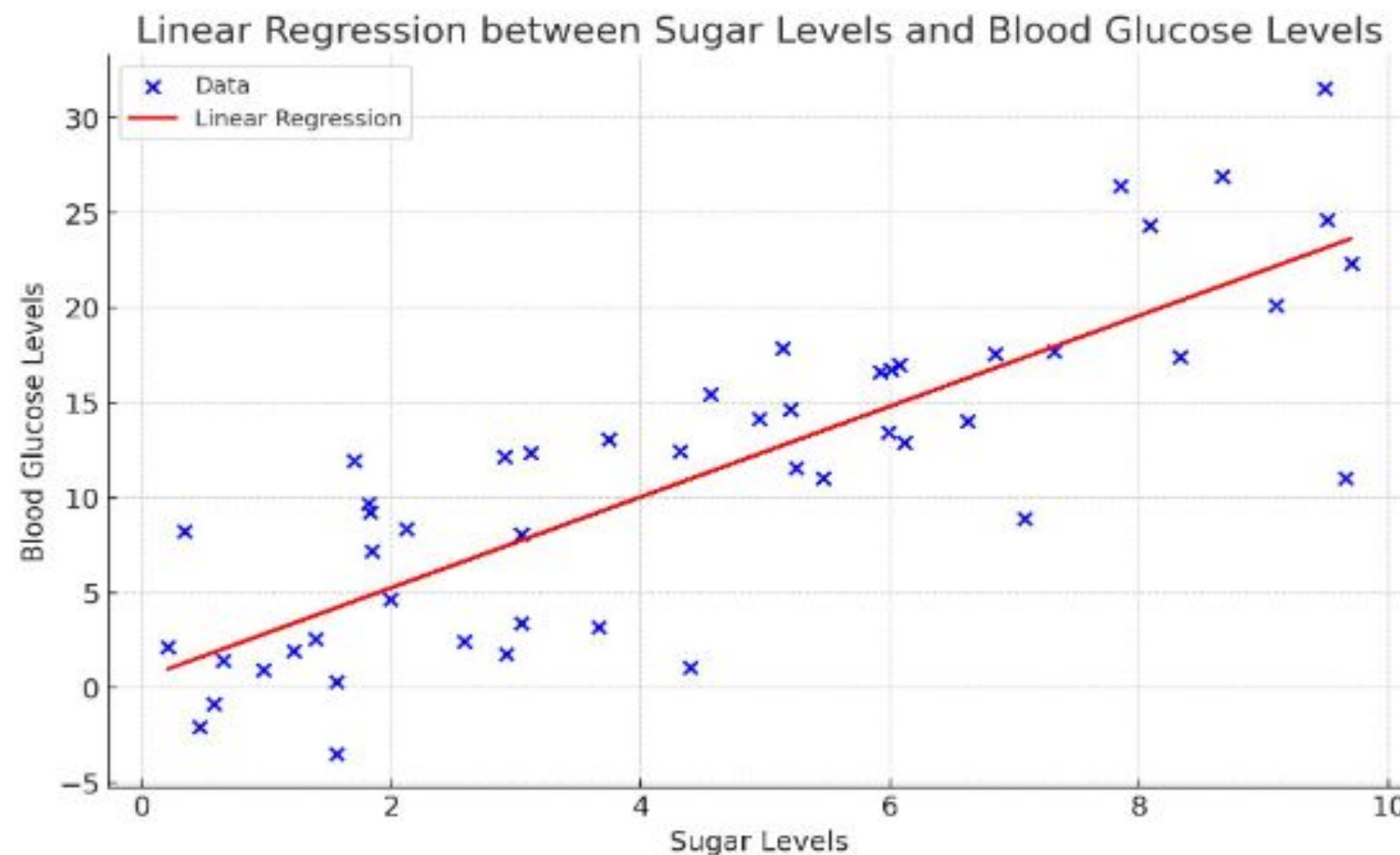
Input (x)	Output (y)
Sugar Levels	Blood Glucose Levels
3.75	13.06
9.51	24.62
7.32	17.72
5.99	13.46
1.56	-3.49

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

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

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

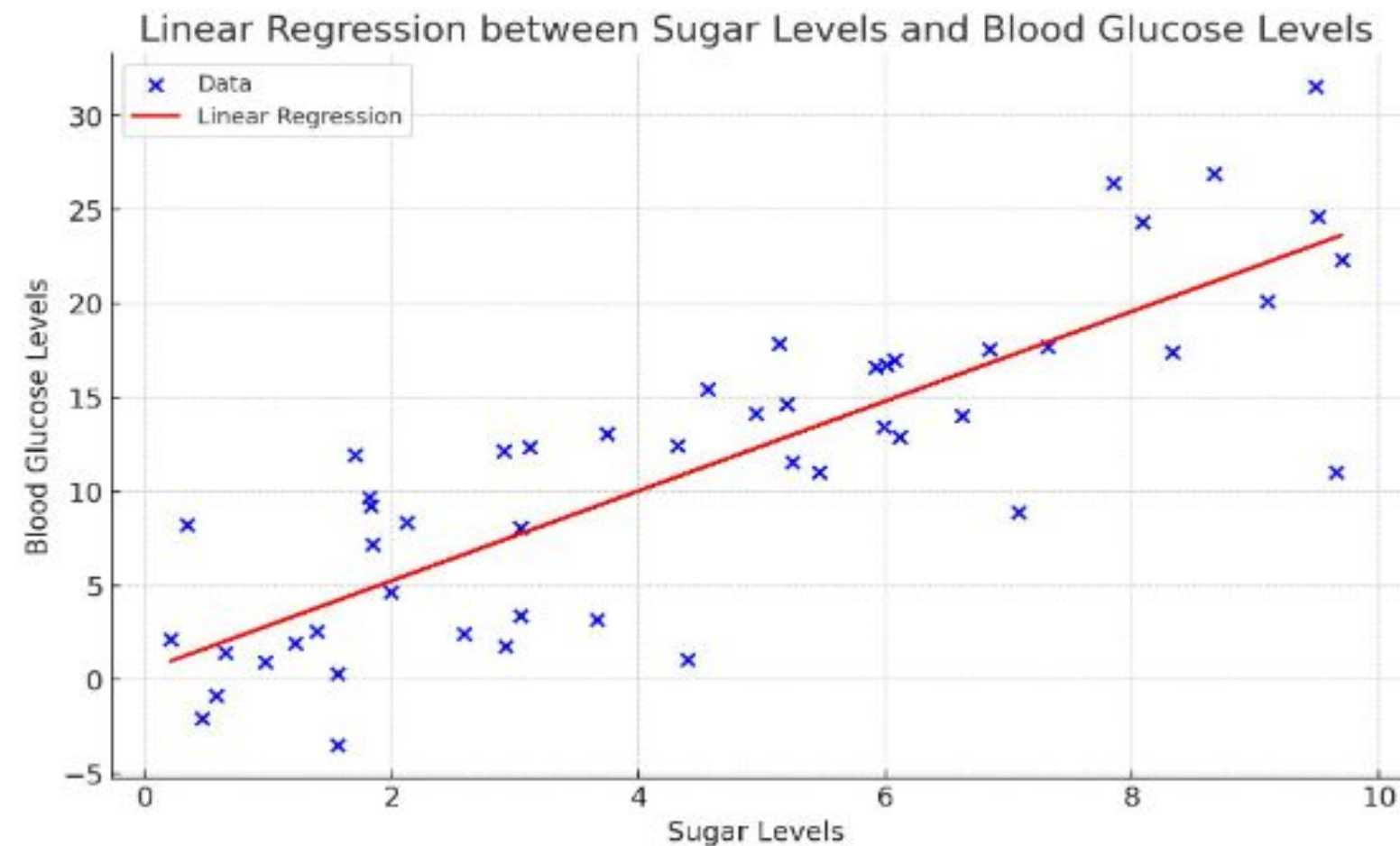
Model

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

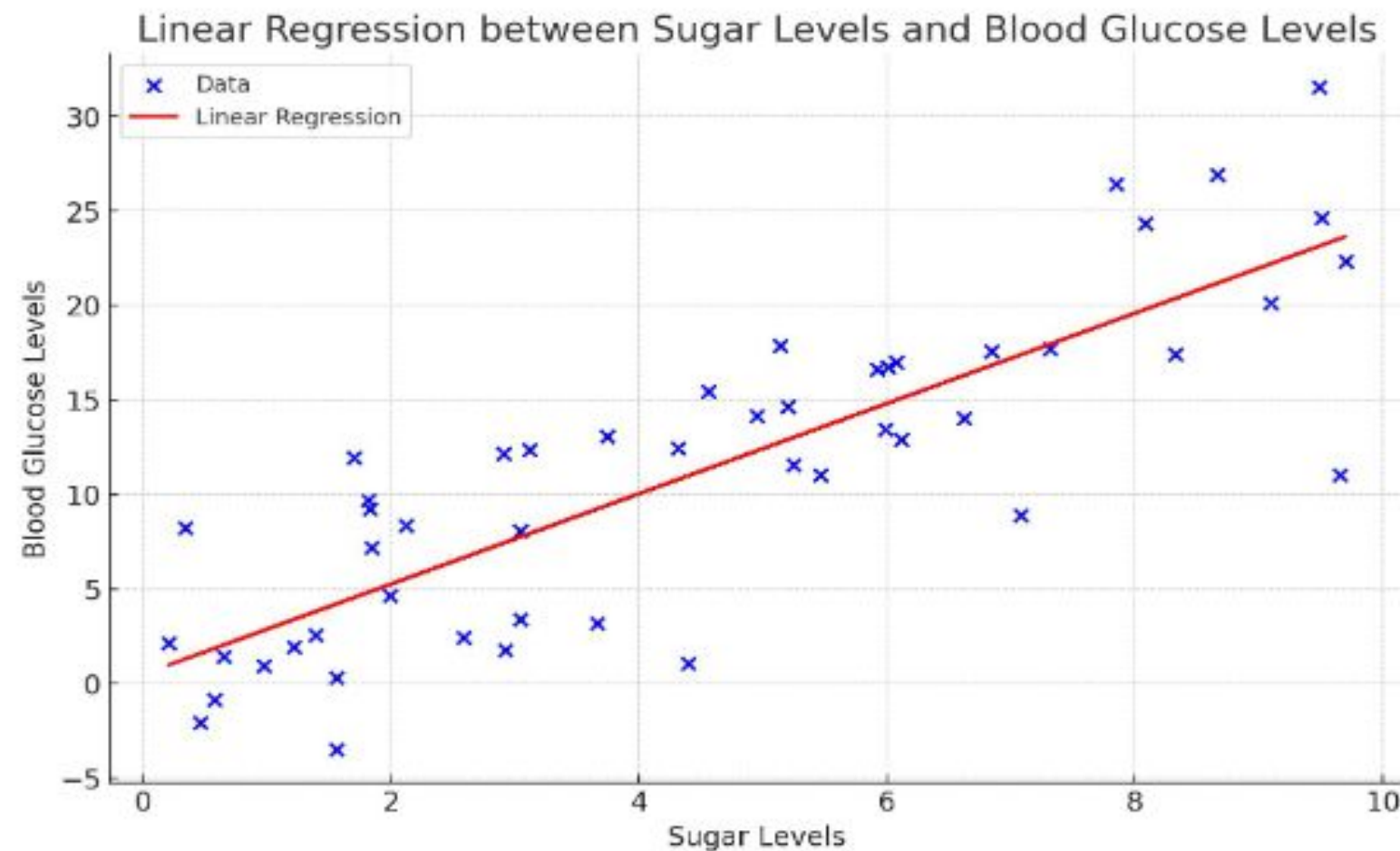
Regression: given point x , predict a numerical value

What do I need in order to predict these outputs?

Error

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

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



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

TRANSFORMATEC



REGRESSION aka Fitting Curves to Data

What do I need in order to predict these outputs?

Least Squares Regression: $y = mx + b$

$$m = \frac{N \sum(xy) - \sum x \sum y}{N \sum(x^2) - (\sum x)^2}$$

$$b = \frac{\sum y - m \sum x}{N}$$

$$y = mx + b$$



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 $d : X \times X \rightarrow \mathbb{R}$ 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



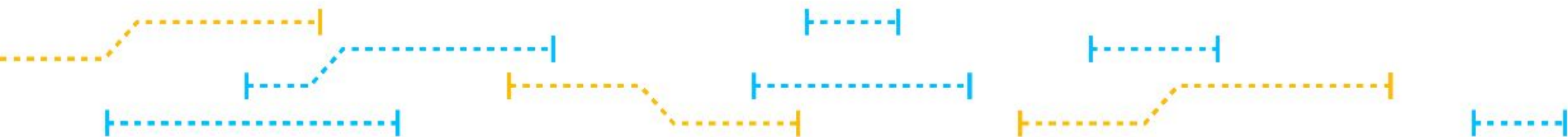
Condition of a Metric

Definition 7.1. A metric d on a set X is a function $d : X \times X \rightarrow \mathbb{R}$ 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

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

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

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

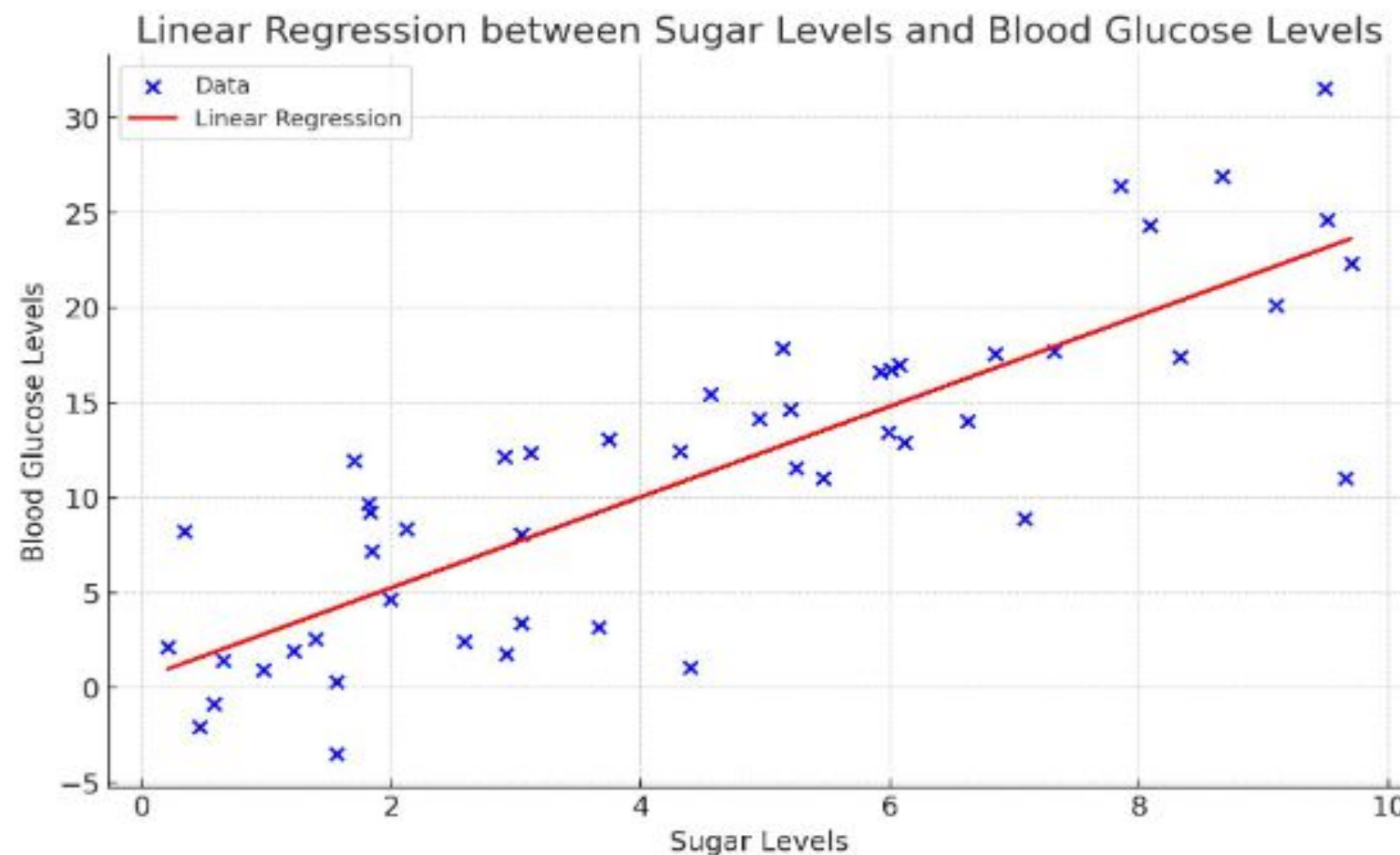
Regression: given point x , predict a numerical value

What do I need in order to predict these outputs?

Error

MSE?

MAE?



Input (x)

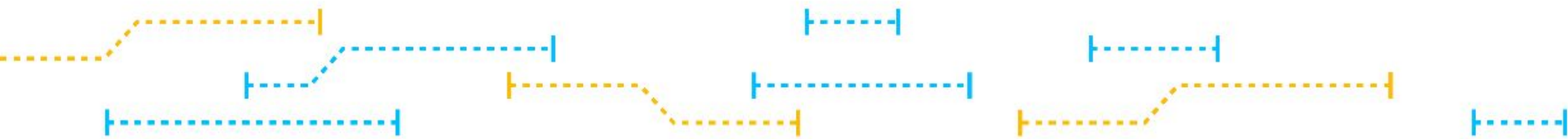
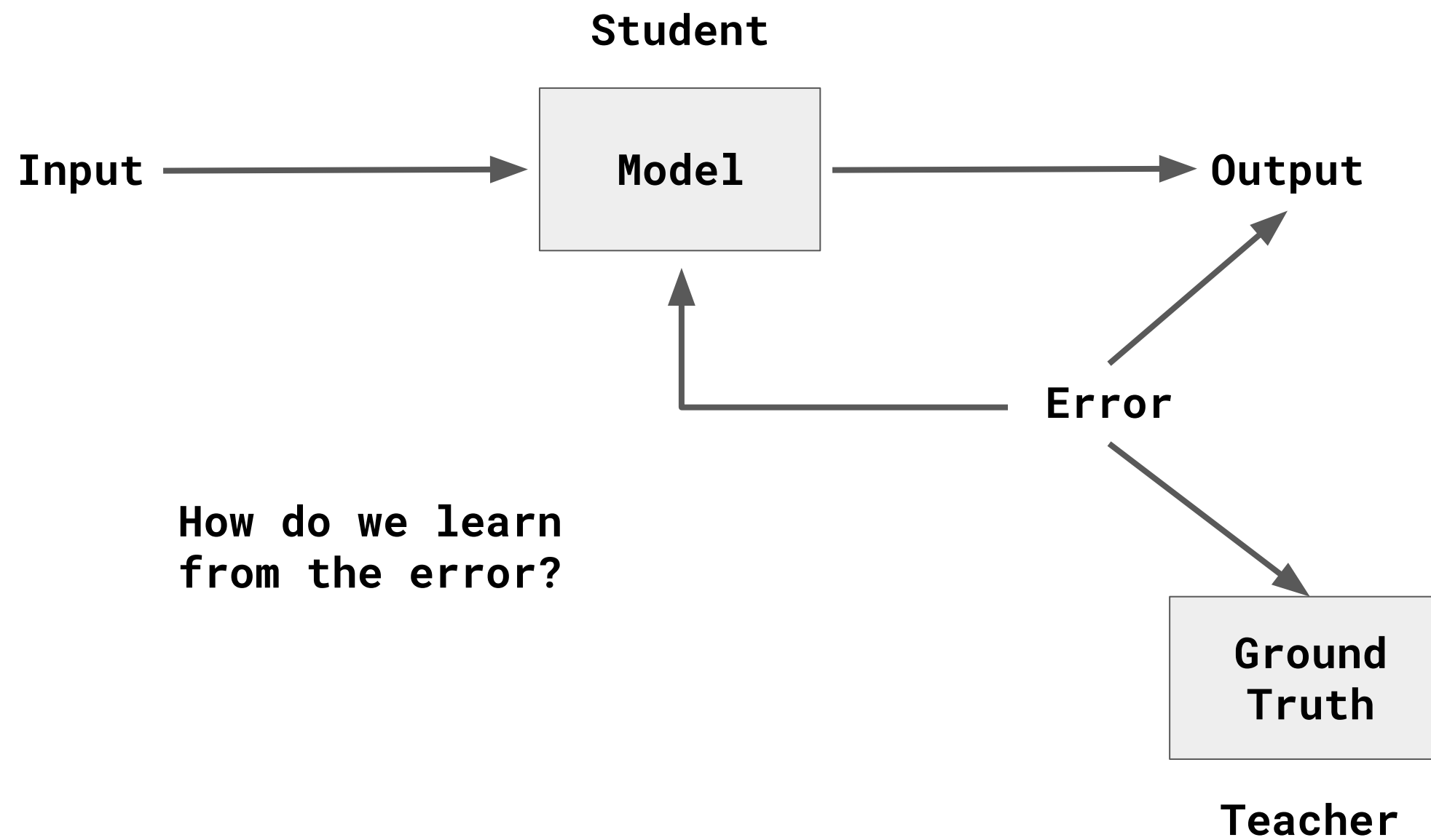
Output (y)

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

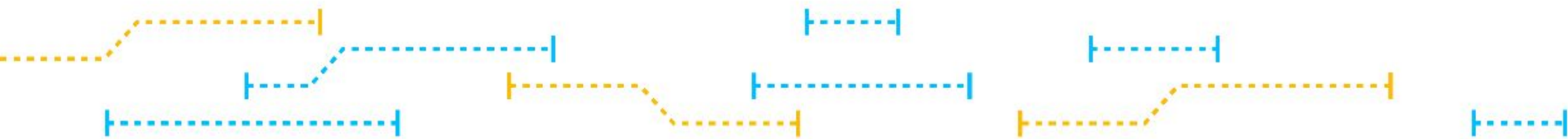
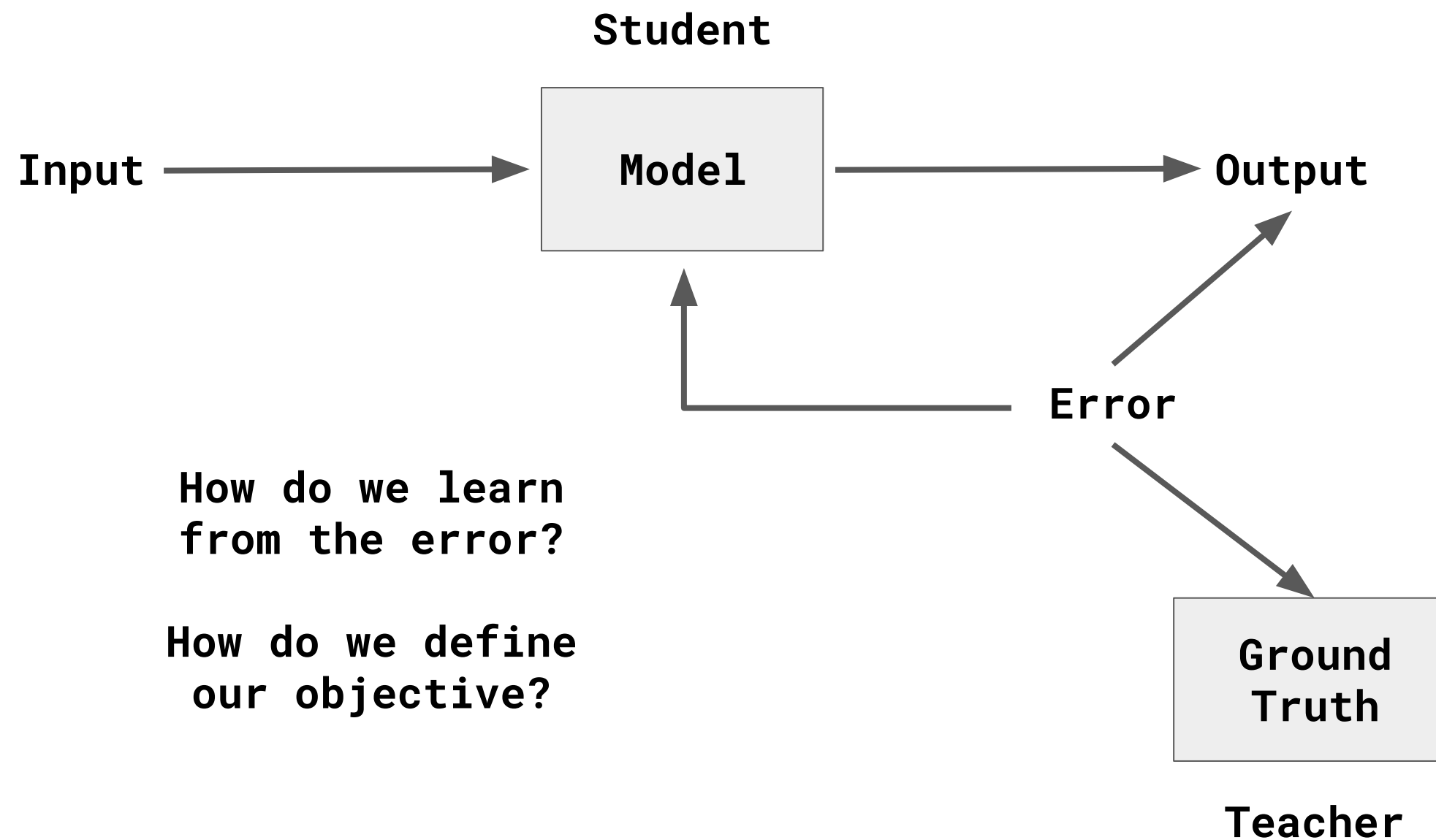
$$f(x) = ax + b \Rightarrow f(x) = wx + b$$

Model

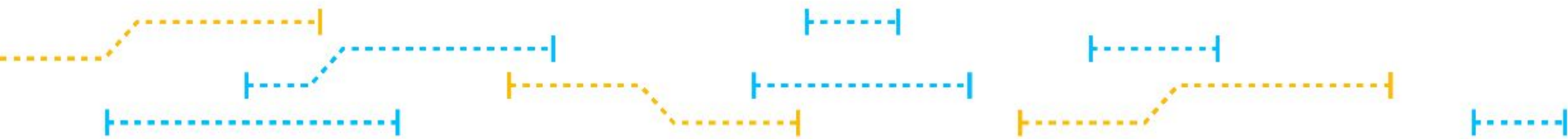
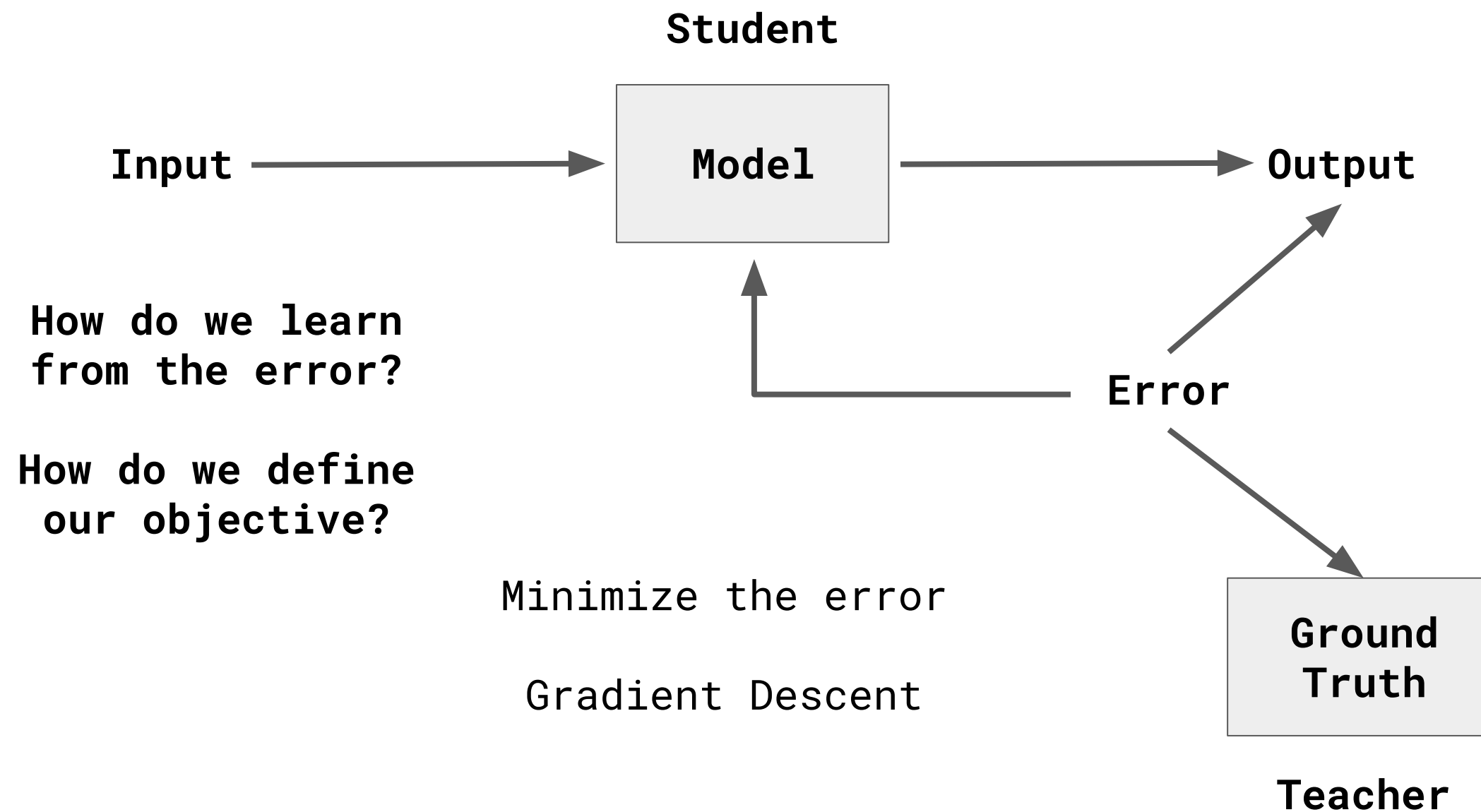
Regression



Regression



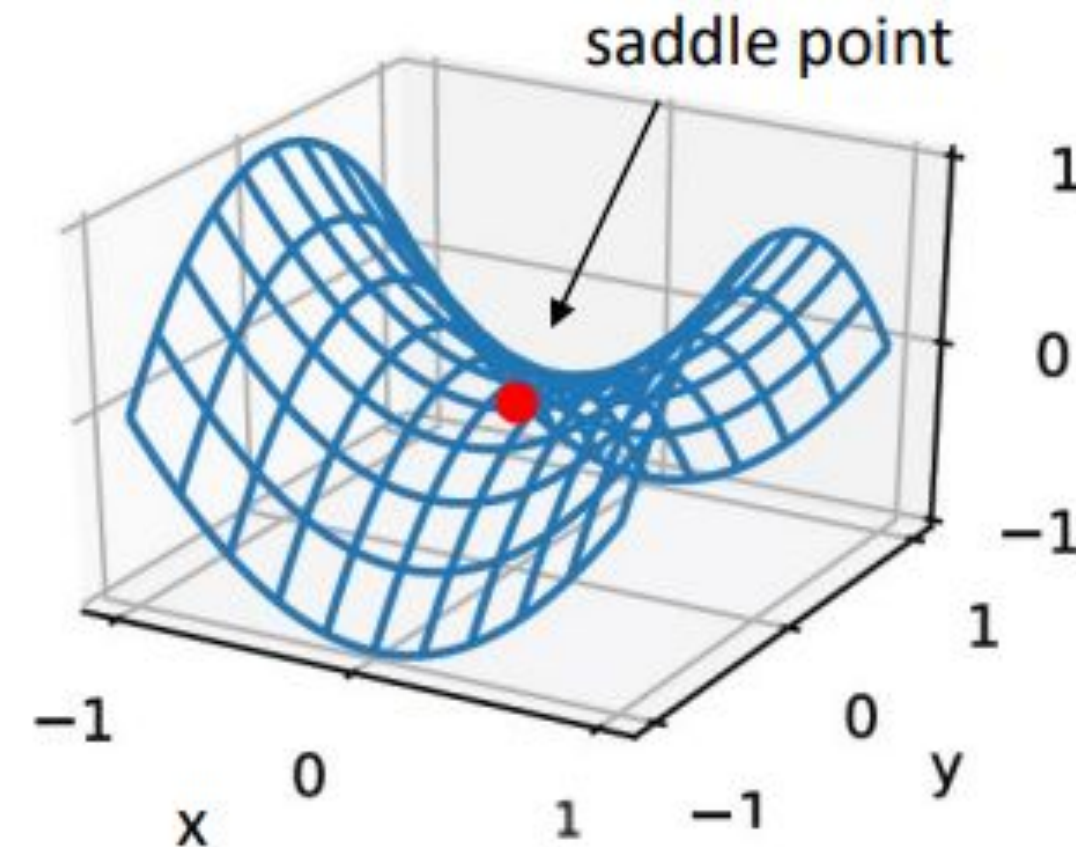
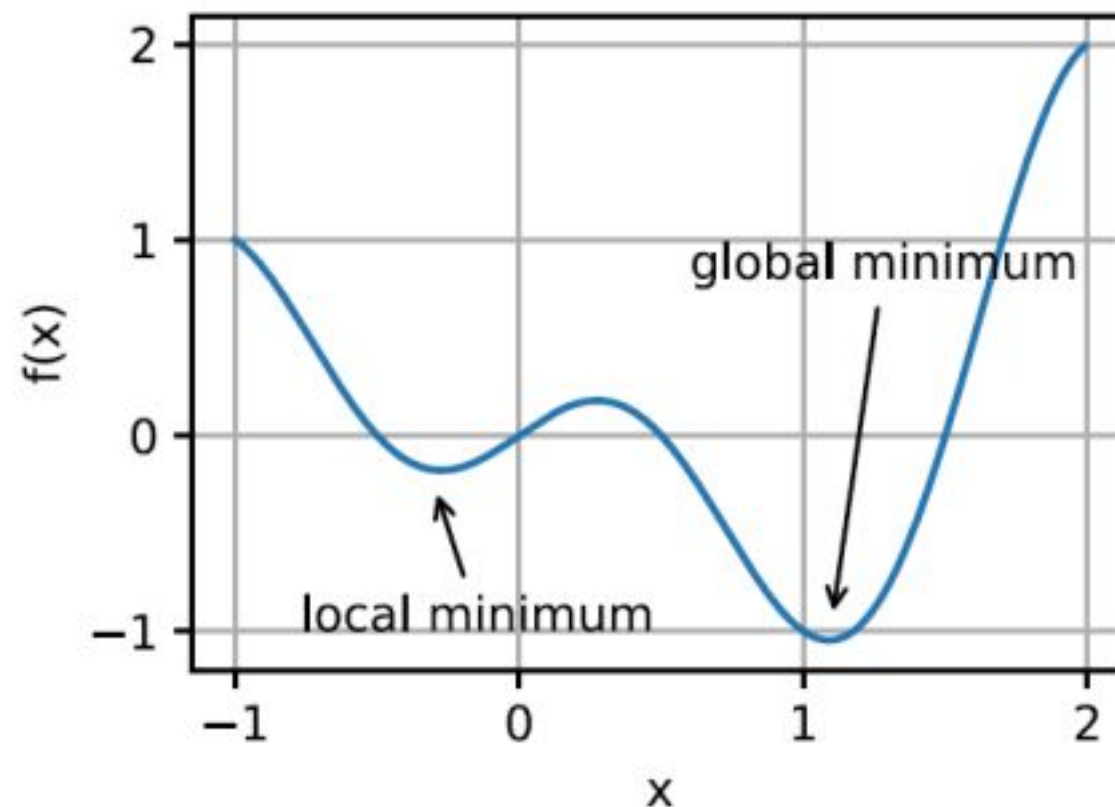
Regression



Gradient Descent - Optimization

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

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

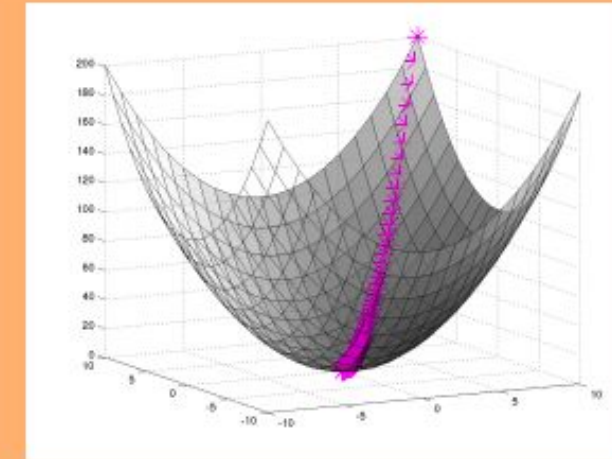


Gradient Descent

Algorithm 1 Gradient Descent

```

1: procedure GD( $\mathcal{D}$ ,  $\theta^{(0)}$ )
2:    $\theta \leftarrow \theta^{(0)}$ 
3:   while not converged do
4:      $\theta \leftarrow \theta - \gamma \nabla_{\theta} J(\theta)$ 
5:   return  $\theta$ 
  
```



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

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

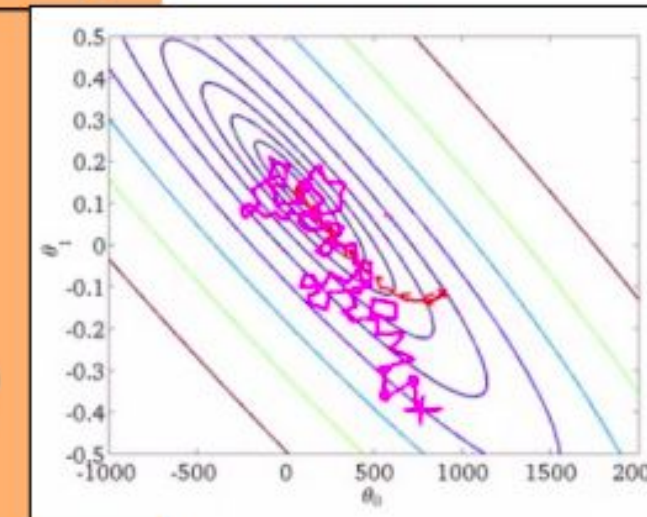


Stochastic Gradient Descent

Algorithm 1 Stochastic Gradient Descent (SGD)

```

1: procedure SGD( $\mathcal{D}, \theta^{(0)}$ )
2:    $\theta \leftarrow \theta^{(0)}$ 
3:   while not converged do
4:     for  $i \in \text{shuffle}(\{1, 2, \dots, N\})$  do
5:        $\theta \leftarrow \theta - \gamma \nabla_{\theta} J^{(i)}(\theta)$ 
6:   return  $\theta$ 
  
```



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

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



Linear Regression

Model Prediction / Output $\longrightarrow f(x) = y' = wx + b$

y

\longleftarrow Ground Truth / Label

Minimize the error (MSE, MAE)

$$\text{Loss} = \text{MSE} = 1/n \sum (f(x) - y)^2$$

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

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



Linear Regression

Minimize the error (MSE, MAE)

$$\text{Loss} = \text{MSE} = 1/2n \sum (f(x) - y)^2$$

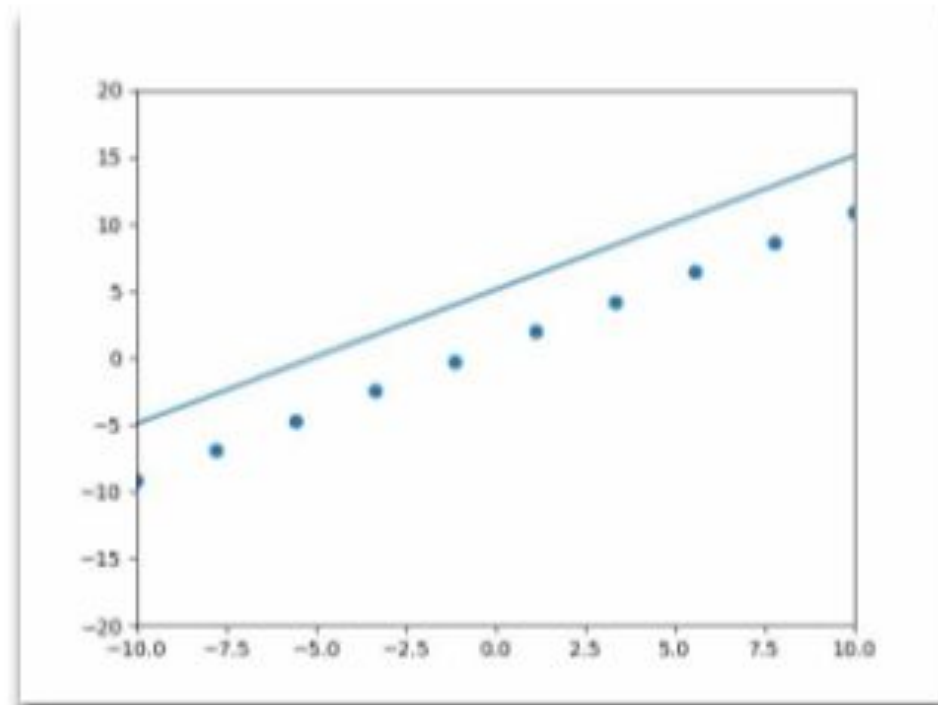
$$w = w - \alpha \frac{\partial \text{Loss}}{\partial w} = w - \alpha \frac{1}{n} \sum (f(x) - y)x$$

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

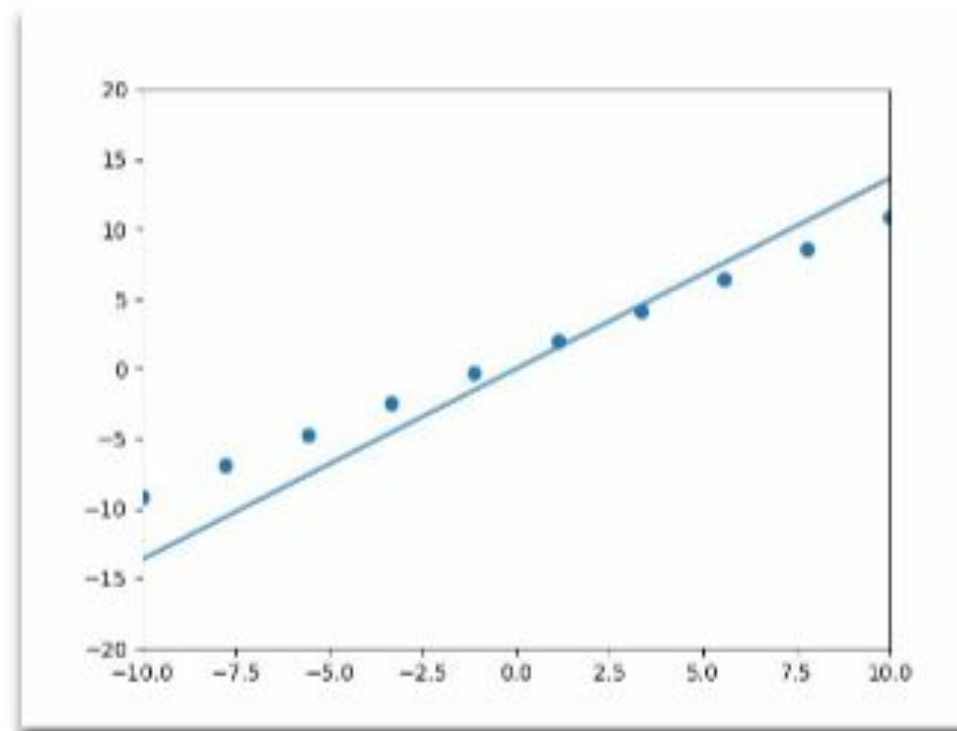


Linear Regression

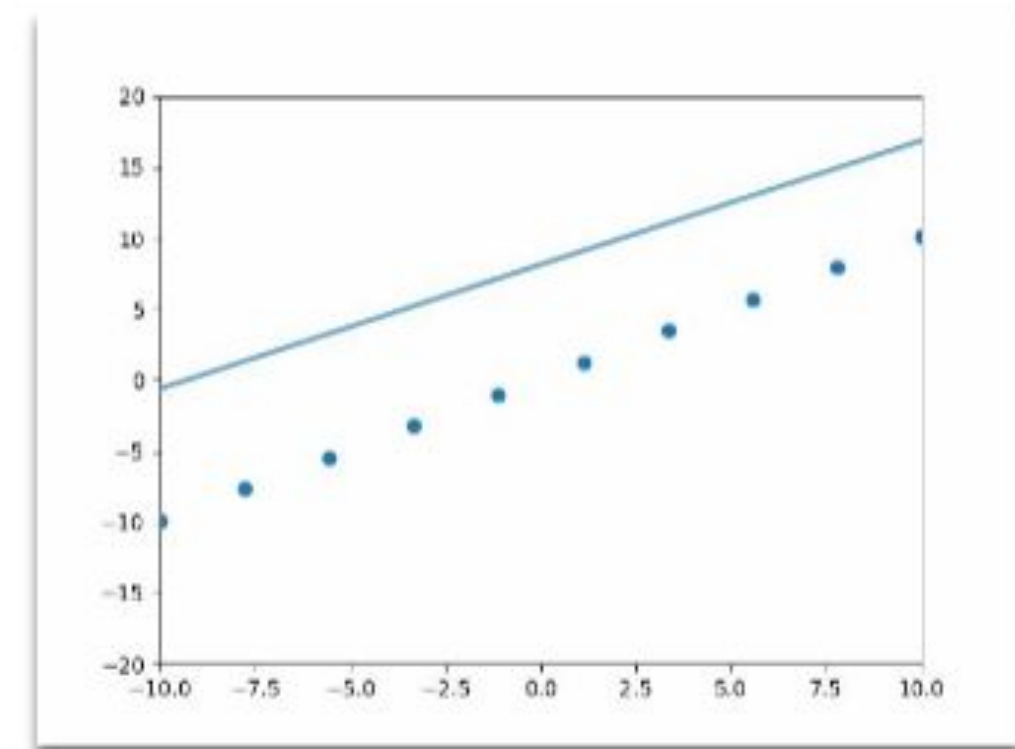
$$L = xw + \boxed{b}$$



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



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



Multidimensional Linear Regression

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

Each house is a data point \mathbf{n} , with observations indexed by \mathbf{j} :

$$\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_j^{(n)}, \dots, x_d^{(n)})$$

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

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

We can then solve for $\mathbf{w} = (w_0, w_1, \dots, w_d)$. 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**:

$$\mathbf{x}^{(n)} = (x_1^{(n)}, \dots, x_j^{(n)}, \dots, x_d^{(n)})$$

We can incorporate the bias 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 $\mathbf{w} = (w_0, w_1, \dots, w_d)$. How?

We can use gradient descent to solve for each **coefficient**, or compute \mathbf{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

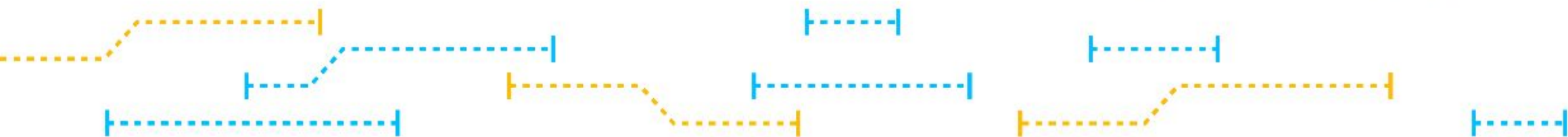
Each house is a data point **n**, with observations indexed by **j**:

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

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

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

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



3.

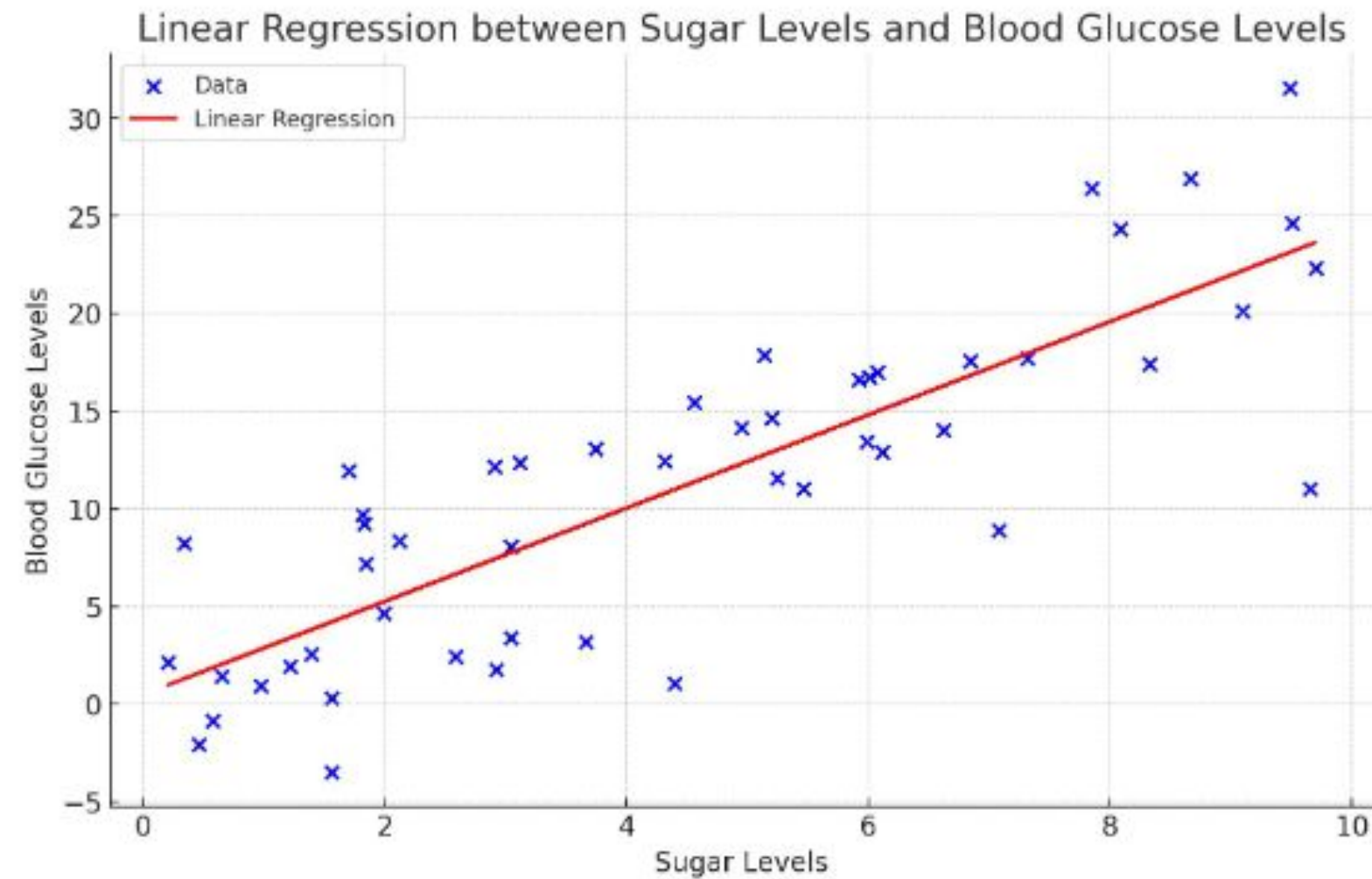


Linear Regression

Example

<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

