LECTURER: TAI LE QUY

MACHINE LEARNING

SUPERVISED LEARNING

TOPIC OUTLINE

UNIT 2

REGRESSION

STUDY GOALS

- Understand the **concept** of regression and when to use it.
- Evaluate a regression model’s **performance**.
- Utilize **regularization** techniques and understand where they are implemented.
- Apply different well-known regression models with the use of **Python**.

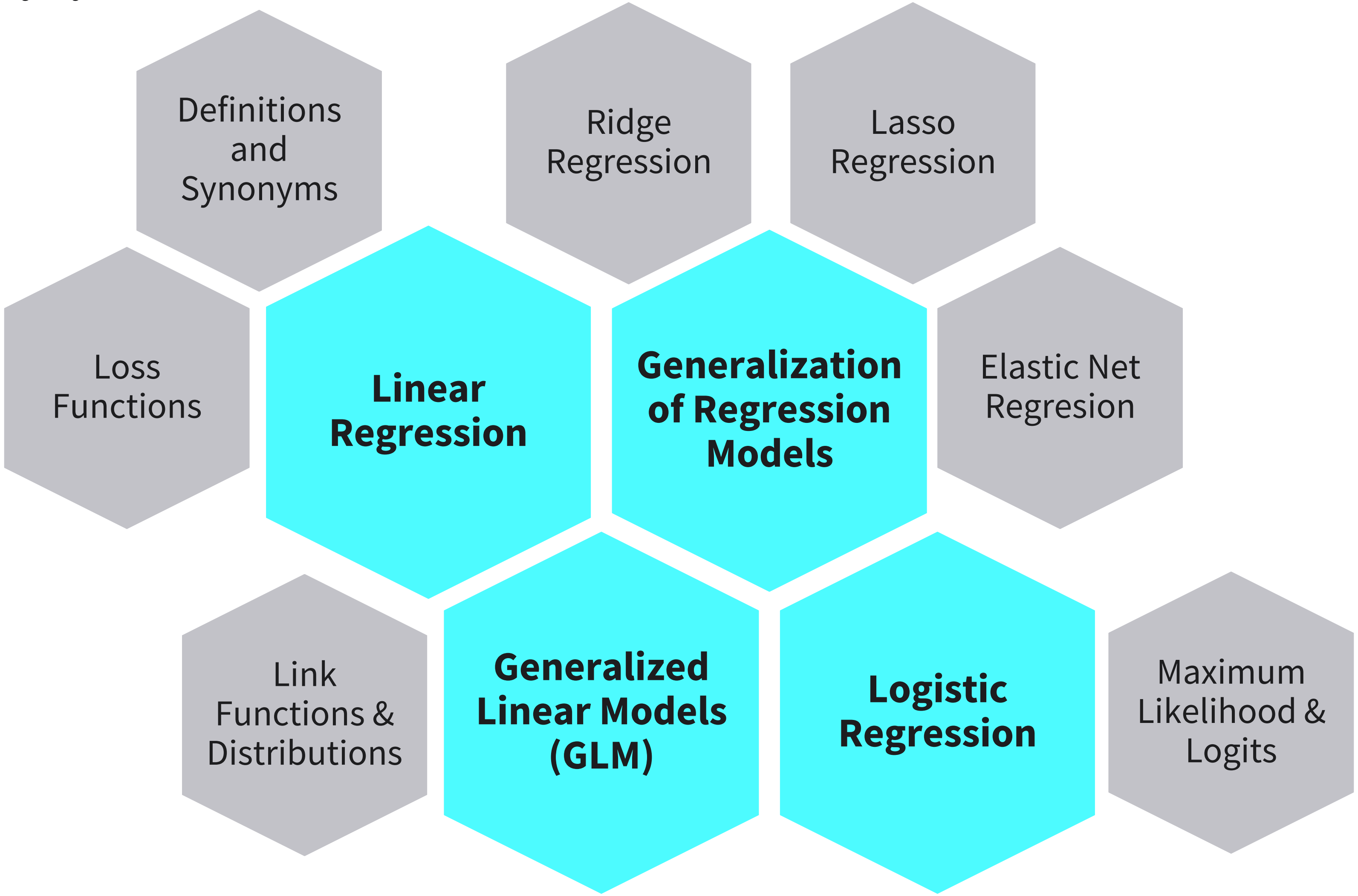
1. Name **three terms** for the **x variables** in linear regression.

2. Explain why we should be alerted if $\mathbf{R^2}$ is **1**.

3. Explain what makes **logistic regression different** from other regression techniques.

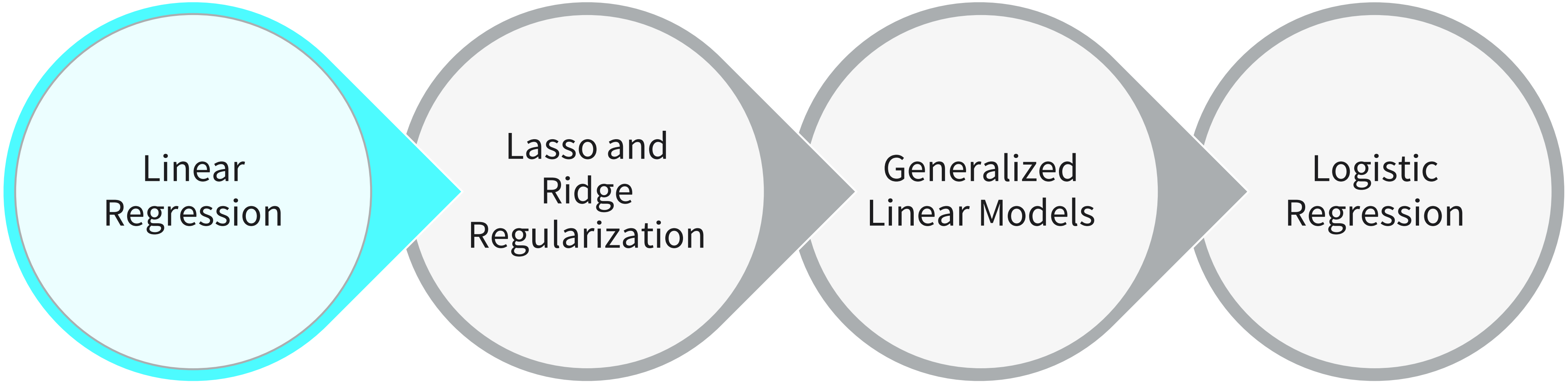
UNIT CONTENT

Img. 1: Regression



Source of image 1: Christian Müller-Kett, 2021.

REGRESSION

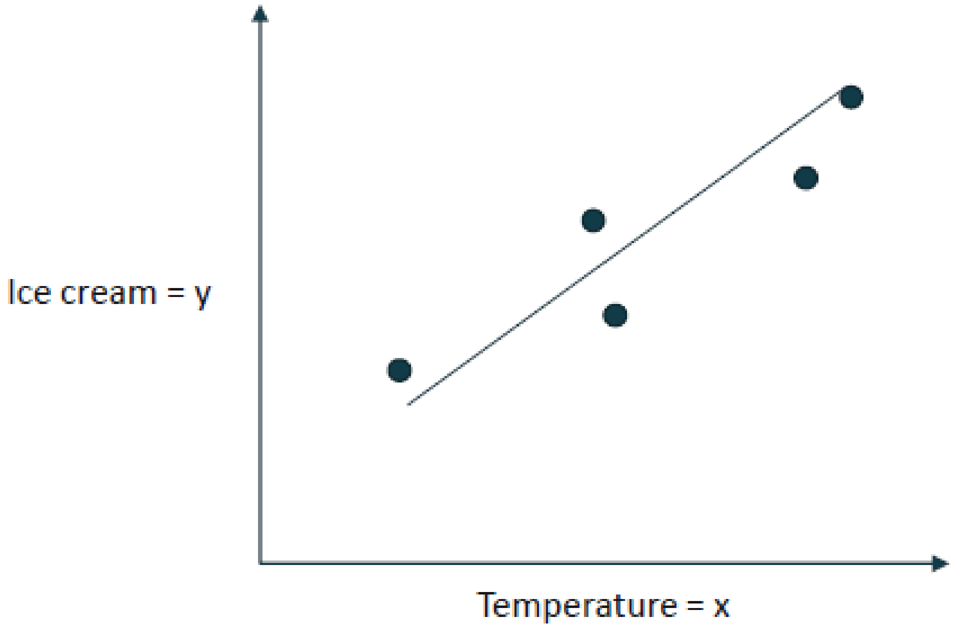

Linear Regression
Lasso and Ridge Regularization
Generalized Linear Models
Logistic Regression

Term	Synonyms
Target variable (Y)	Label, y variable, dependent variable
Input (X)	X variables, independent variables, predictors, features
Coefficient (ω_i)	Weight, slope, regression coefficient
(y-axis) Intercept (ω_0)	Bias
Loss function	Cost function, target function, objective function, error function

Important Regression Terms and Their Synonyms

Source of image: Course book

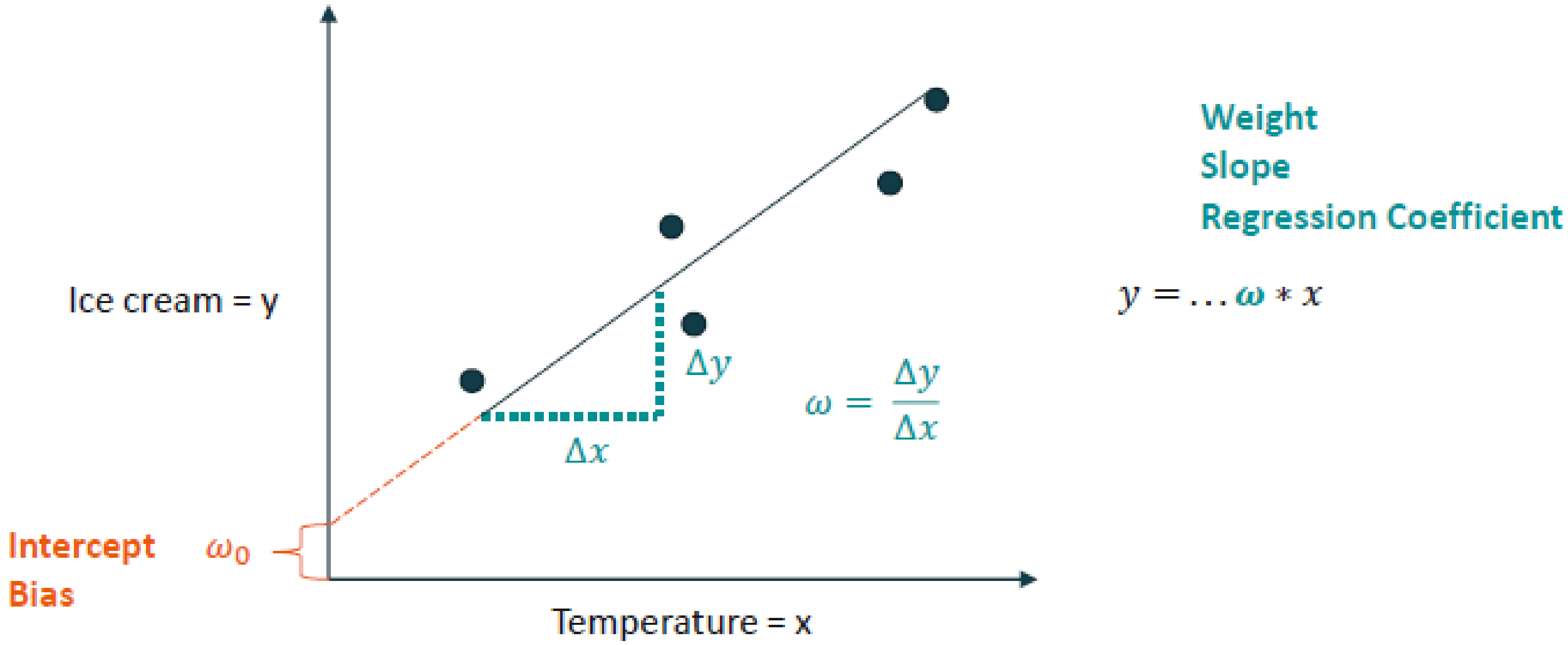
LINEAR REGRESSION



$$y = ? * x$$

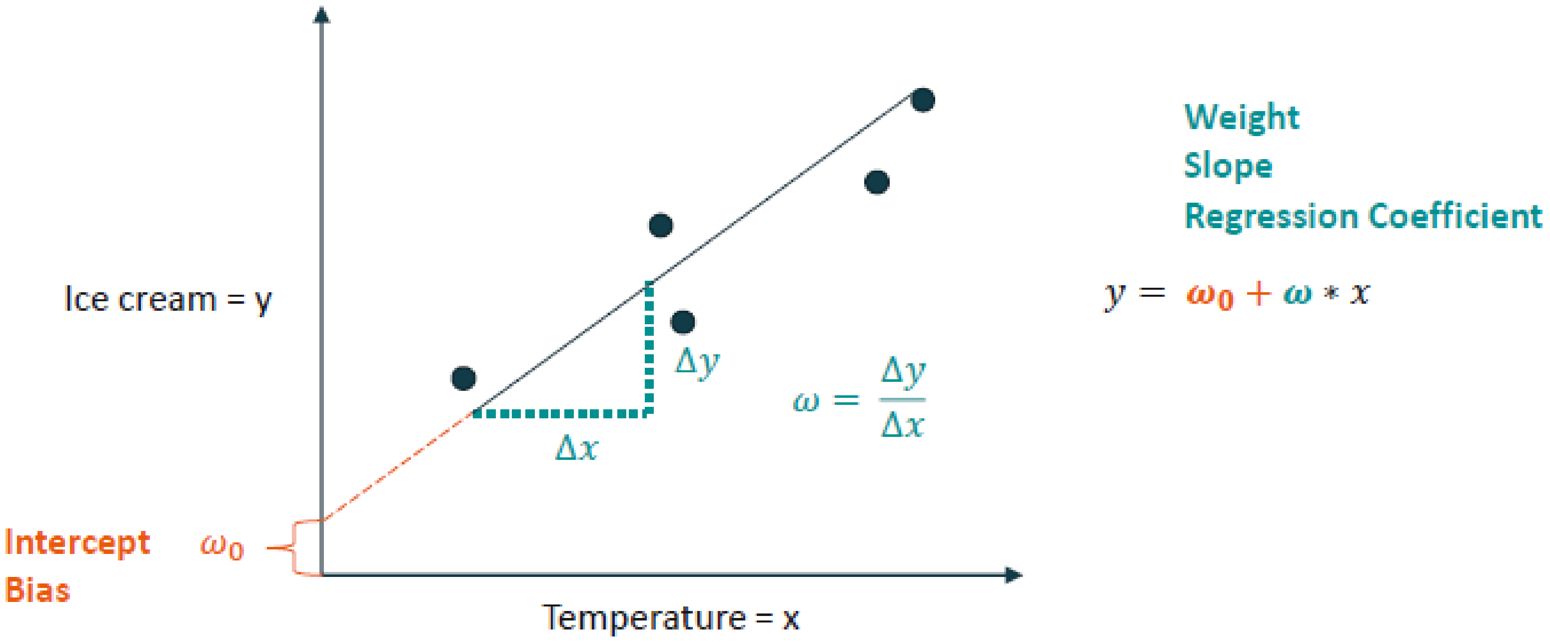
Source of image: https://learn.iu.org/

LINEAR REGRESSION



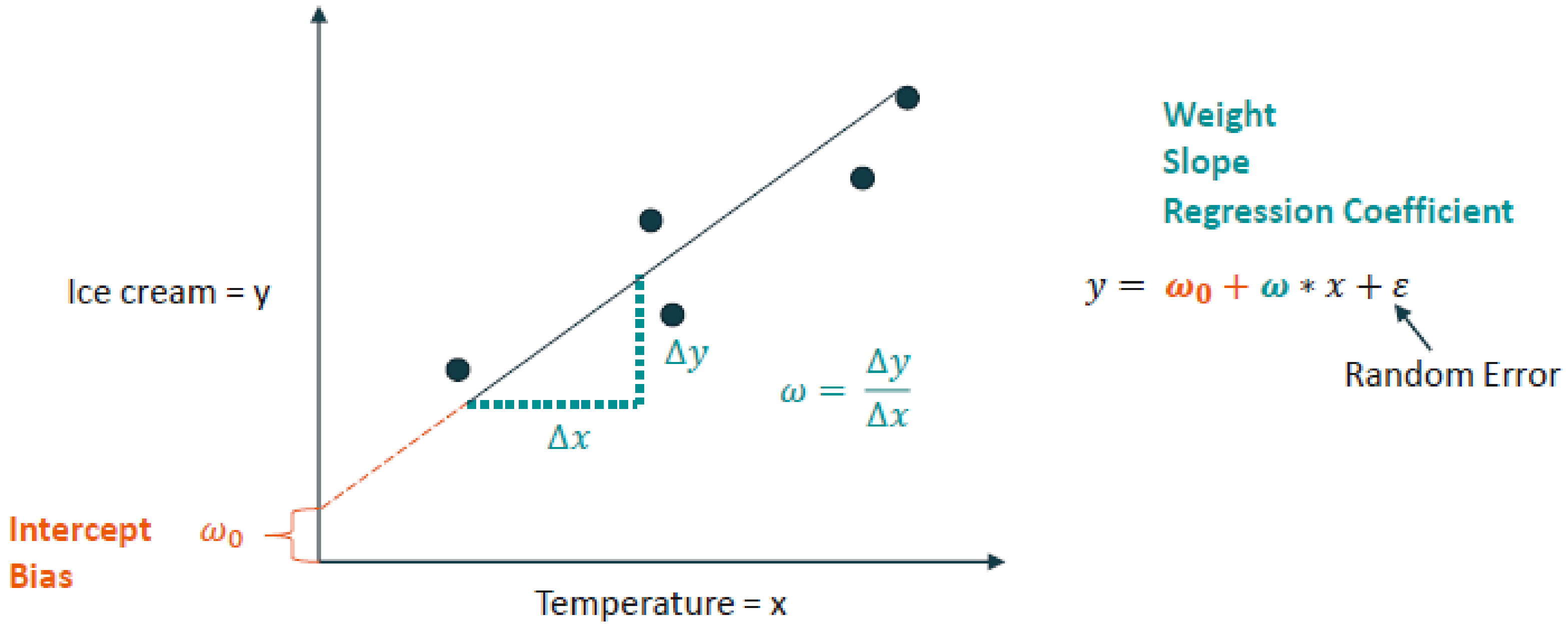
Source of image: https://learn.iu.org/

LINEAR REGRESSION



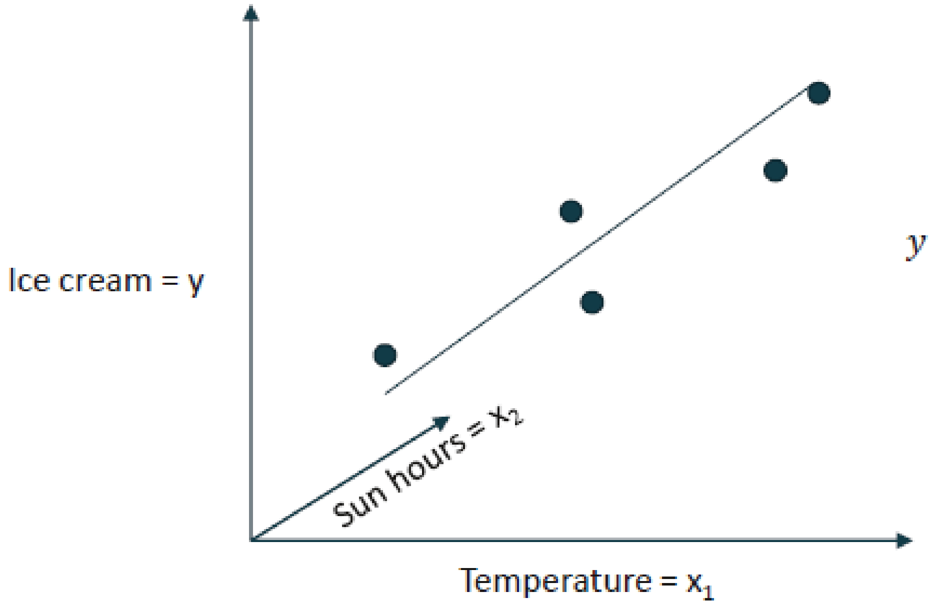
Source of image: https://learn.iu.org/

LINEAR REGRESSION



Source of image: https://learn.iu.org/

LINEAR REGRESSION

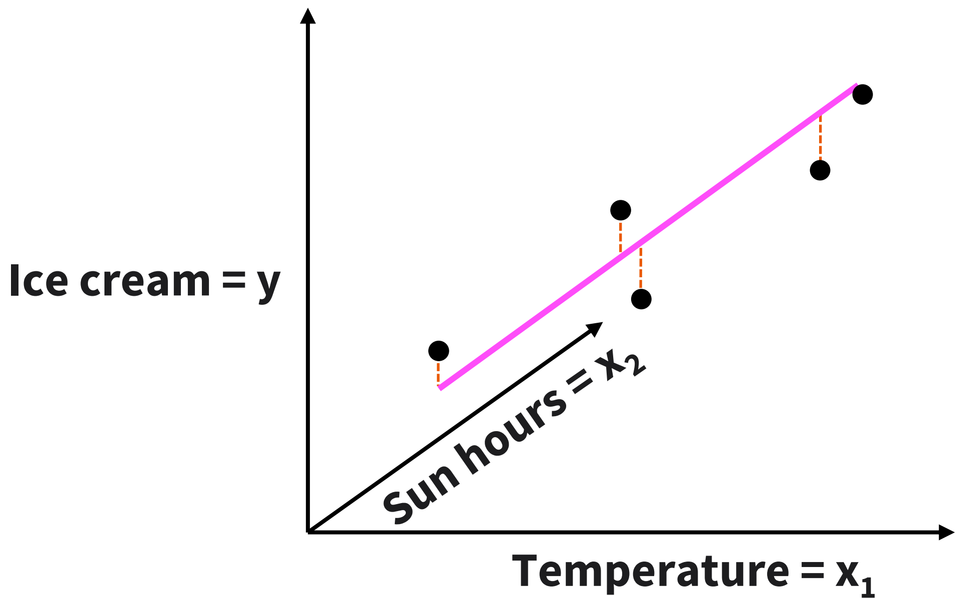


$$y = \omega_0 + \omega_1 * x_1 + \omega_2 * x_2 + \varepsilon$$

$$y = \omega_0 + \sum_{i=1}^{n} \omega_i * x_i + \varepsilon$$

Source of image: https://learn.iu.org/

Img. 2: Linear regression



$$y = \omega_0 + \sum_{i=1}^{n} \omega_i * x_i + \varepsilon$$

$$Residual = R_i = y_i - \hat{y}_i$$

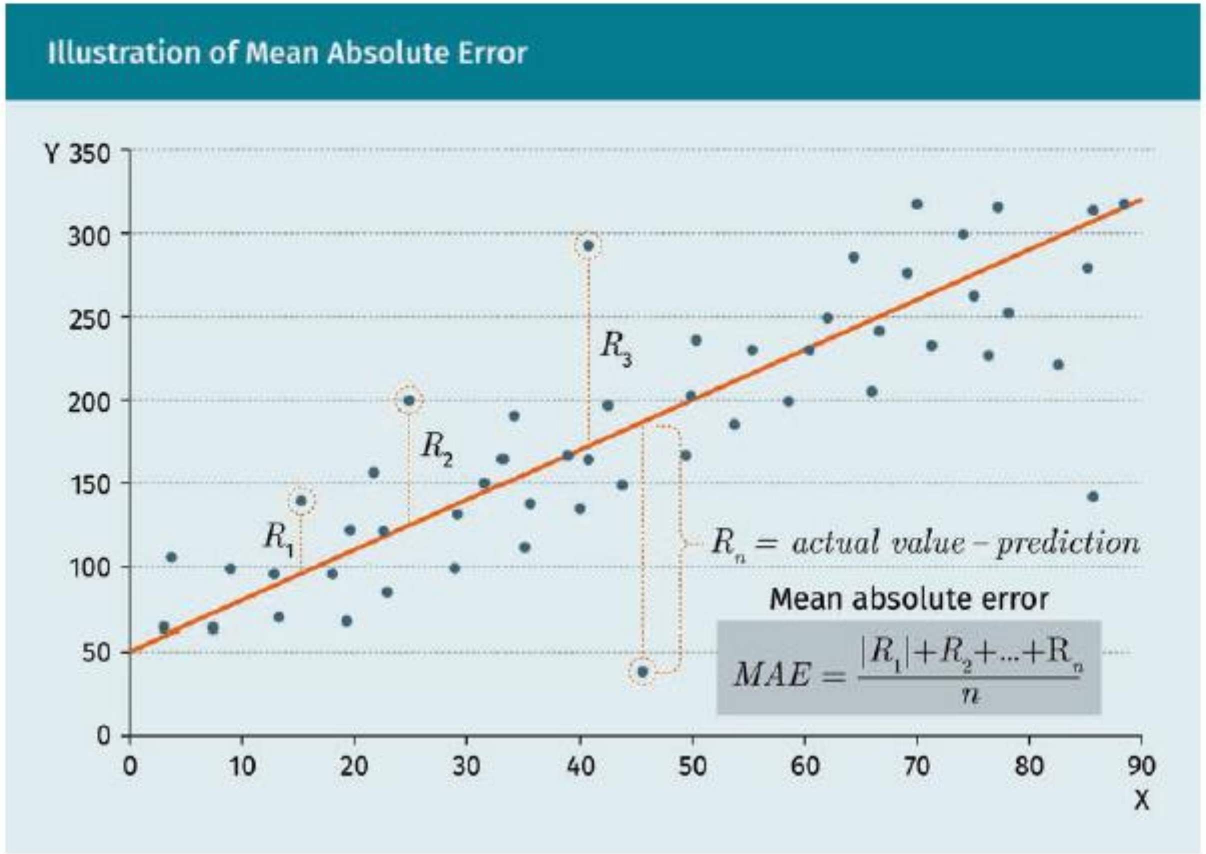
$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{RSS}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

Coefficient of determination

Source of image 2: Christian Müller-Kett, 2021.

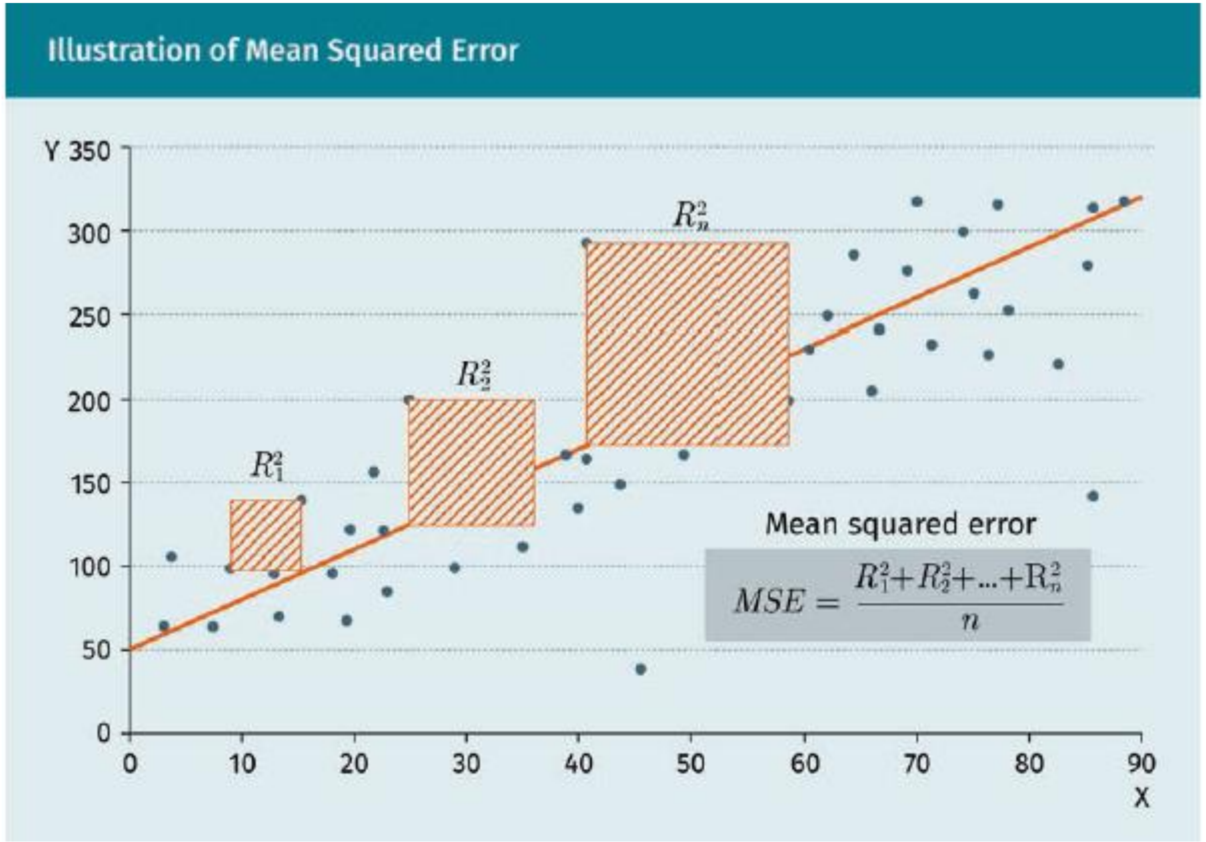
Mean absolute error (MAE): absolute difference between all predictions and the actual values



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

Source of image: Course book

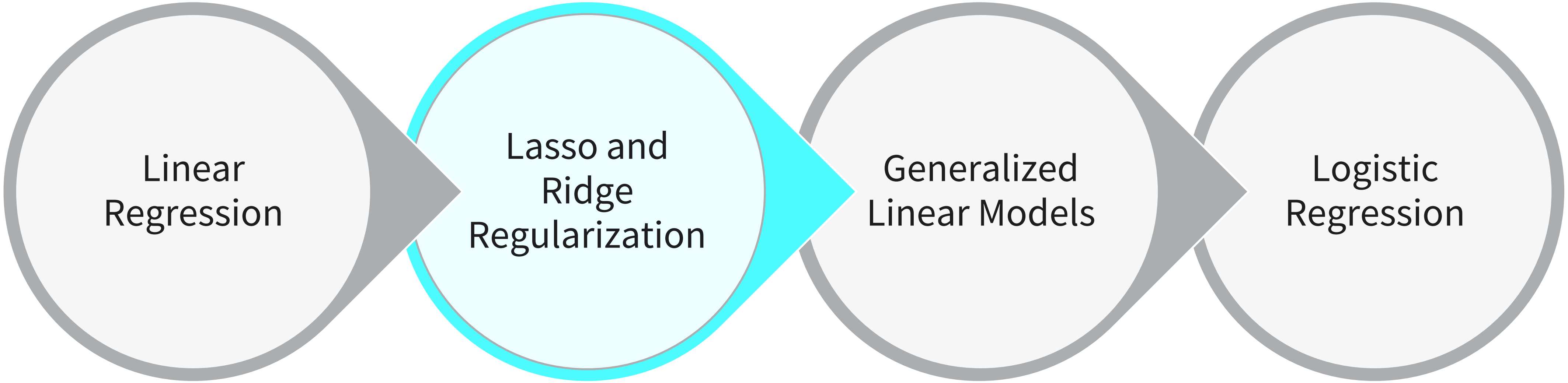
- Mean squared error (MSE): The deviations between the actual and the predicted values are taken to the square.
- Root mean square error (RMSE): the result is telegraphed in the unit of the label being predicted



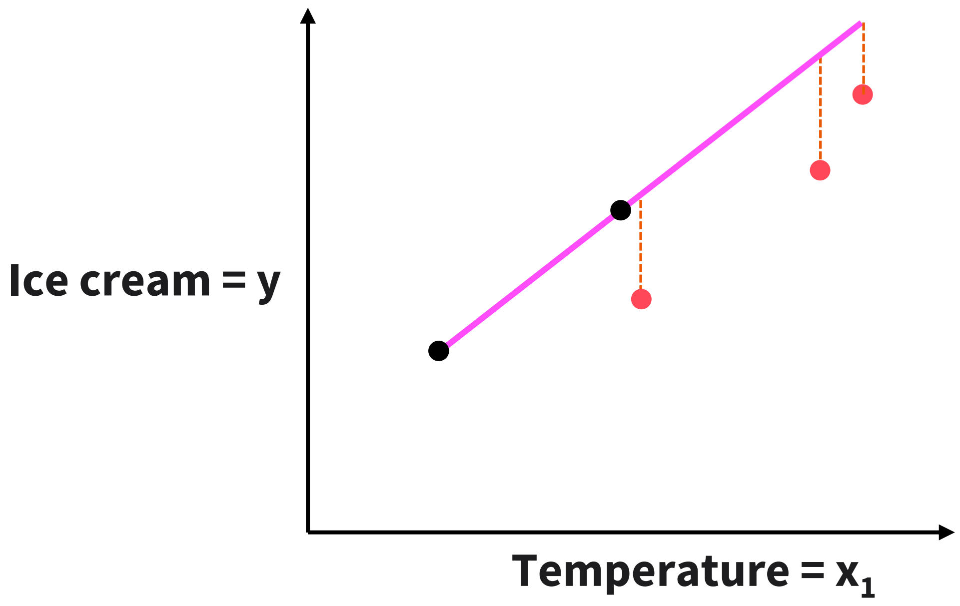
$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \qquad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}$$

Source of image: Course book

REGRESSION


Linear Regression
Lasso and Ridge Regularization
Generalized Linear Models
Logistic Regression

Img. 3: Generalization



$$y = \omega_0 + \sum_{i=1}^{n} \omega_i * x_i + \varepsilon$$

Ridge regression

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boldsymbol{\alpha} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\omega_i^2}$$

The penalty size, also known as the L2 **norm**

Source of image 3: Christian Müller-Kett, 2021.

Ridge regression

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boldsymbol{\alpha} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\omega_i^2}$$

Lasso regression

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boldsymbol{\alpha} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\boldsymbol{\omega_i}|$$

Elastic net regression

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \boldsymbol{r} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\omega_i^2} + (\boldsymbol{1} - \boldsymbol{r}) \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} |\boldsymbol{\omega_i}|$$

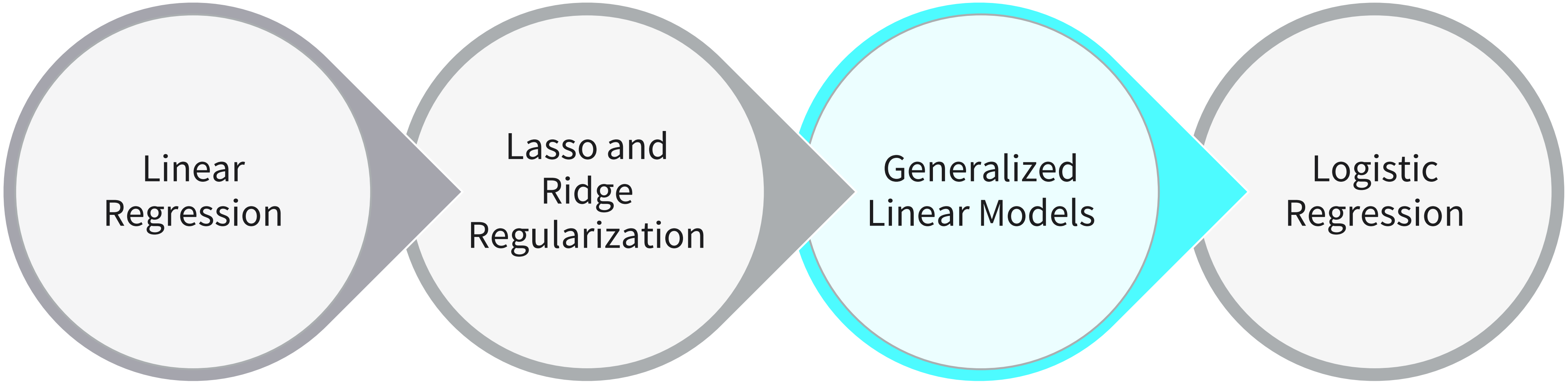
- Lasso (least absolute shrinkage and selection operator)
  - It differs from ridge regression only in that the L2 norm is exchanged for the L1 norm
- The loss function:

$$RSS + \alpha \sum_{i=1}^{n} |\omega_i|$$

- Elastic net: selects important coefficients, as does lasso regression, and is effective in handling correlated features, as is ridge regression.
  - r=0: lasso regression
  - r=1: ridge regression

$$\alpha \sum_{i=1}^{n} \left(r\omega_i^2 + (1-r)|\omega_i| \right)$$

REGRESSION


Linear Regression
Lasso and Ridge Regularization
Generalized Linear Models
Logistic Regression

Assumptions of linear regression models

- **Linear** relationship
- **Normally distributed** residuals
- **Homoscedasticity** (constant variance)

- These assumptions are **relaxed** in GLMs
- Assume that values of **Y** are **drawn from a distribution.**
- Predicting the **mean** of this distribution
- Introducing a **link function**

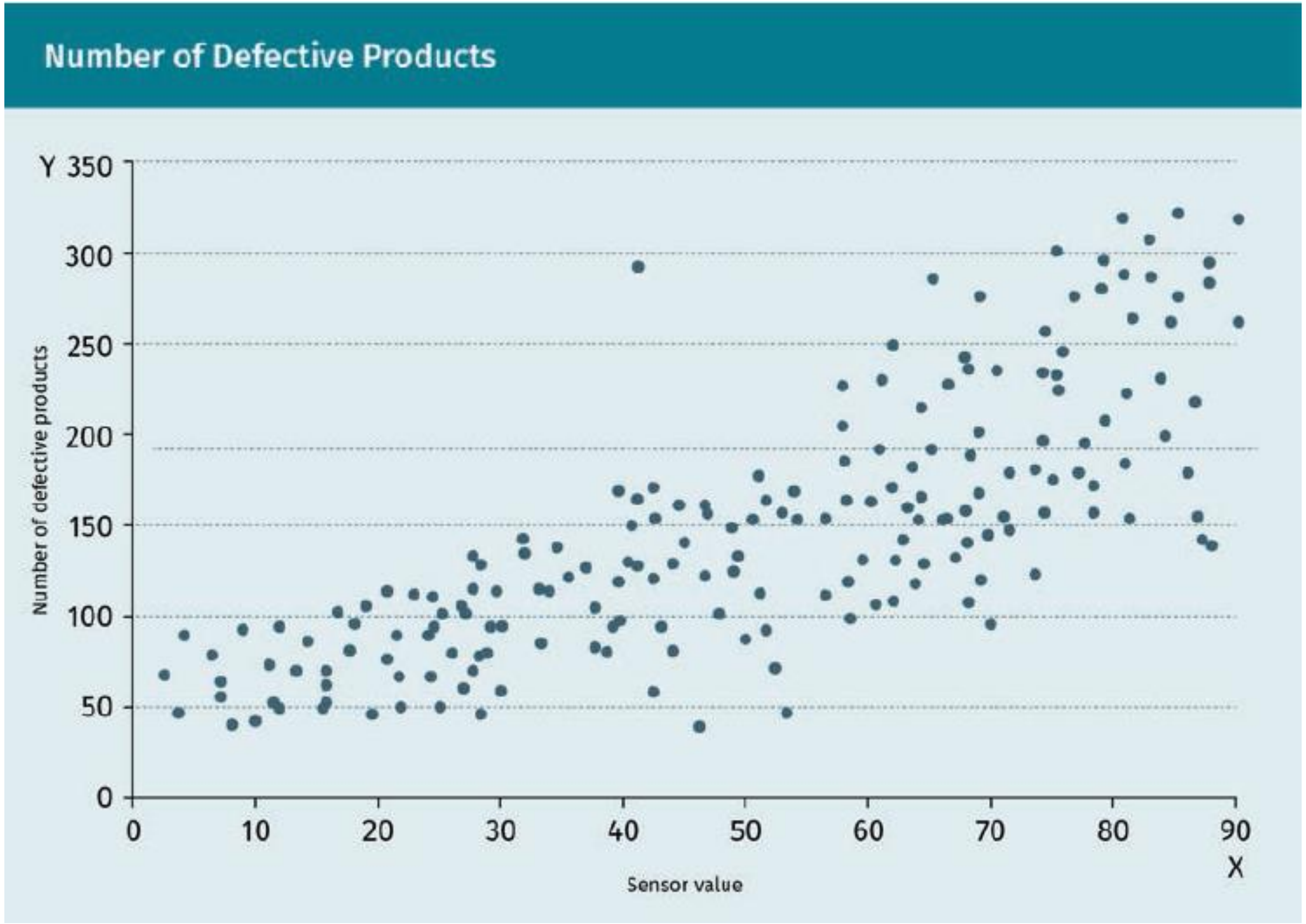
- Generalized linear models (GLMs) are a category of expanded linear regression models.
- General models are developed by relaxing the assumptions of linear models.
- GLMs all essentially comprise the following three components:

1. A linear predictor $\eta_i = \omega_0 + \omega_1 x_{1i} + \ldots + \omega_p x_{pi}$
2. A probability distribution that generates the target variable $\mathbf{Y}$
3. A monotone differentiable **link function** $g(\mu_i) = \eta_i$ describing how the mean depends on the linear predictor $\eta_{i.}$

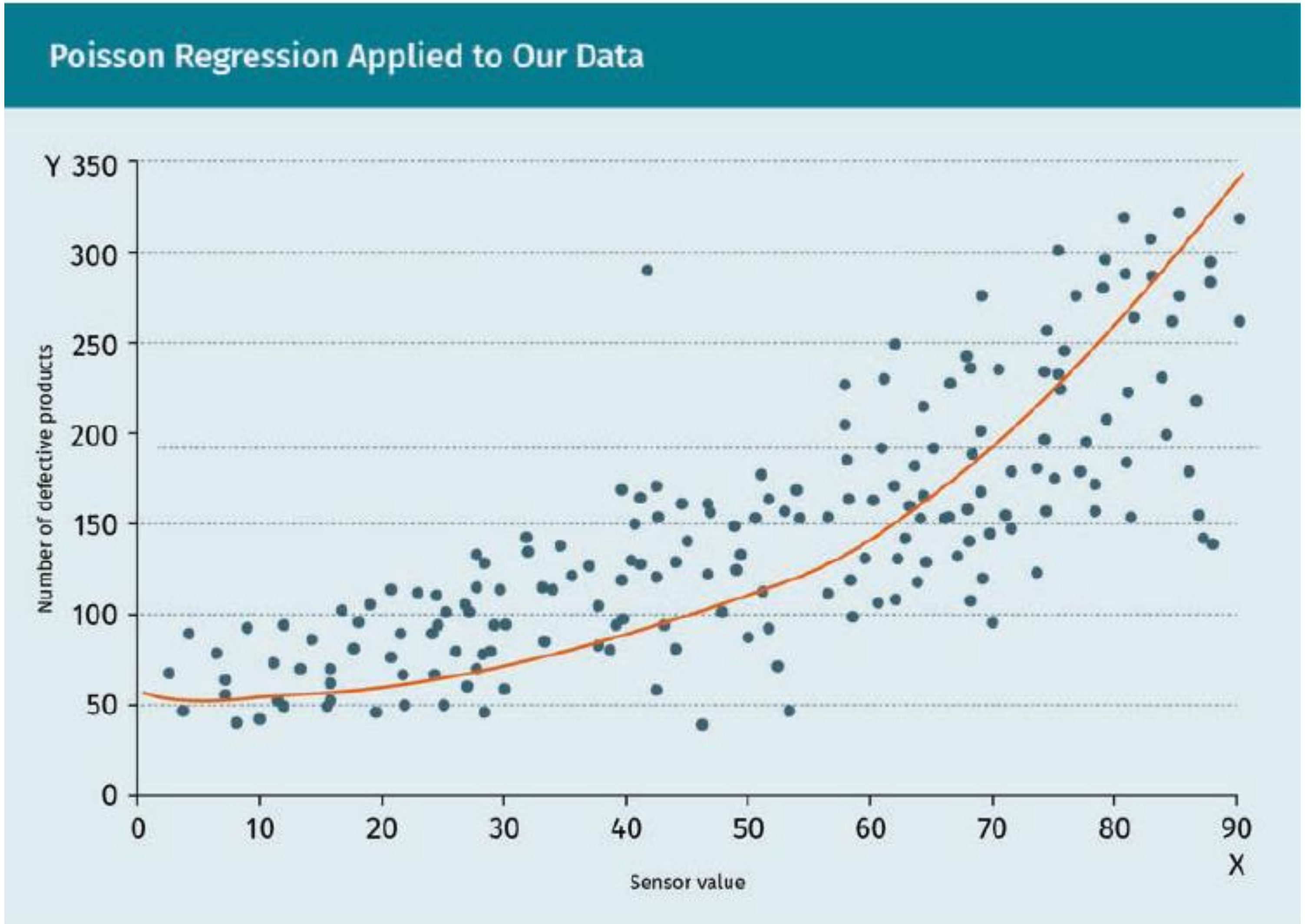
Distribution	Use	Notation	Link function
Gaussian	Linear repose	N(μ,σ2)	"Identity": μ
Poisson	Counts of events	N(μ)	Log(μ)
Bernoulli	Outcome of single yes/no occurrences	Bern(p)	Logit(μ)
Binomial	Count of yes occurren-ces out of n yes/no events	Bin(n, μ)/n	Logit(μ)

Some Probability Distributions and Their Link Functions

GENERALIZED LINEAR MODELS



Source of image: Course book

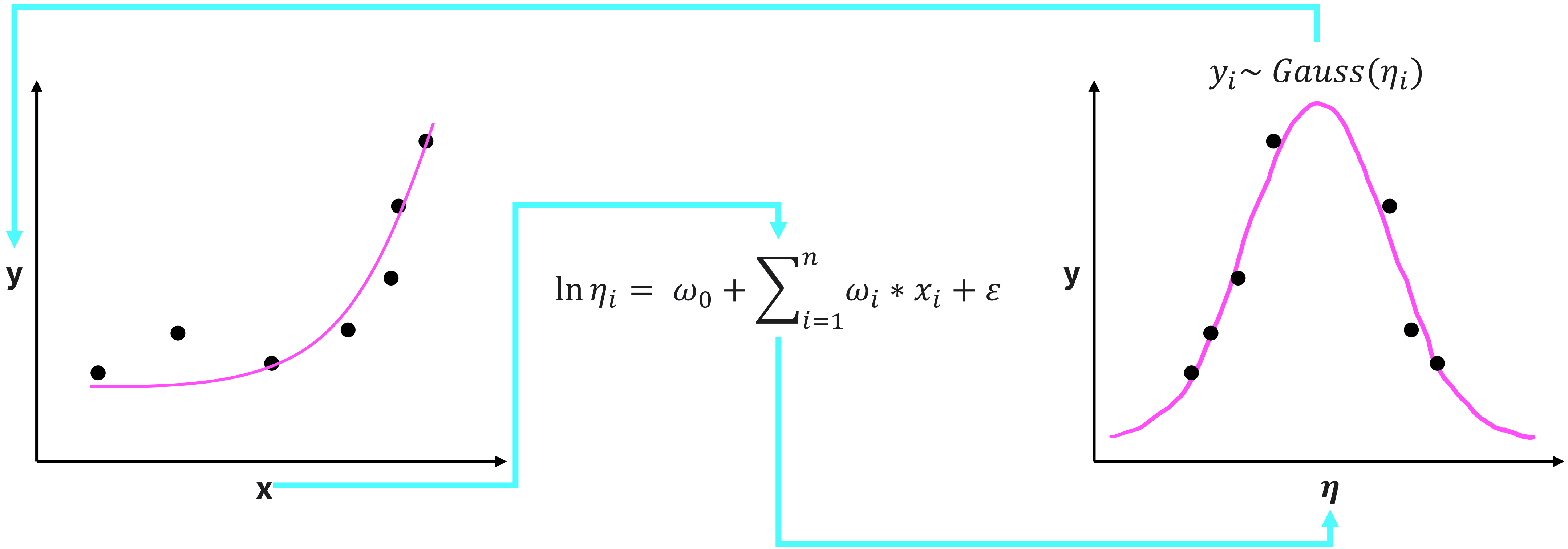


Poisson regression

$$ln\ \eta_i = \omega_0 + \omega_1 x_{1i}$$
$$y_i \sim Poisson(\eta_i)$$

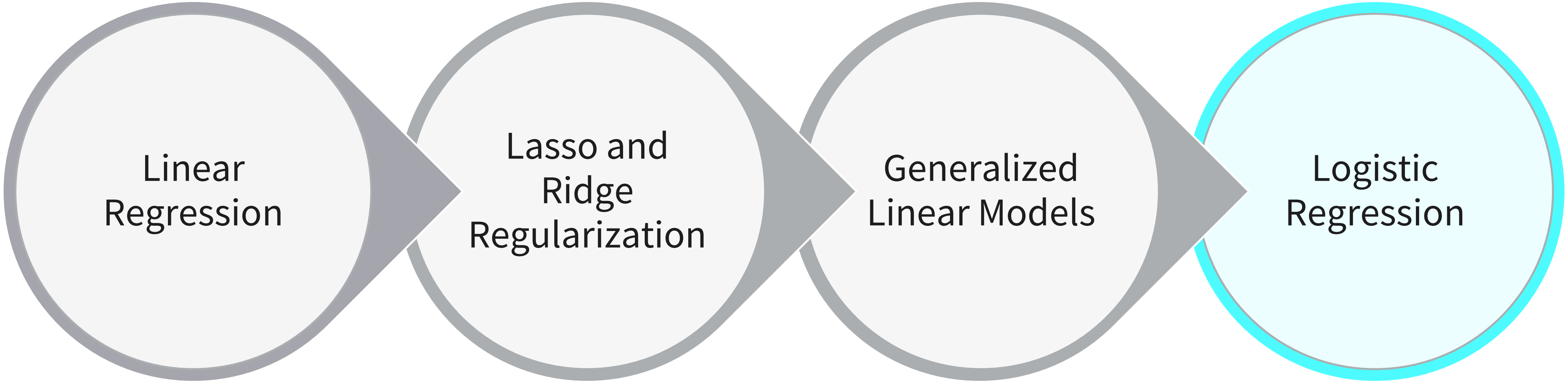
GENERALIZED LINEAR MODELS

Img. 4: GLM

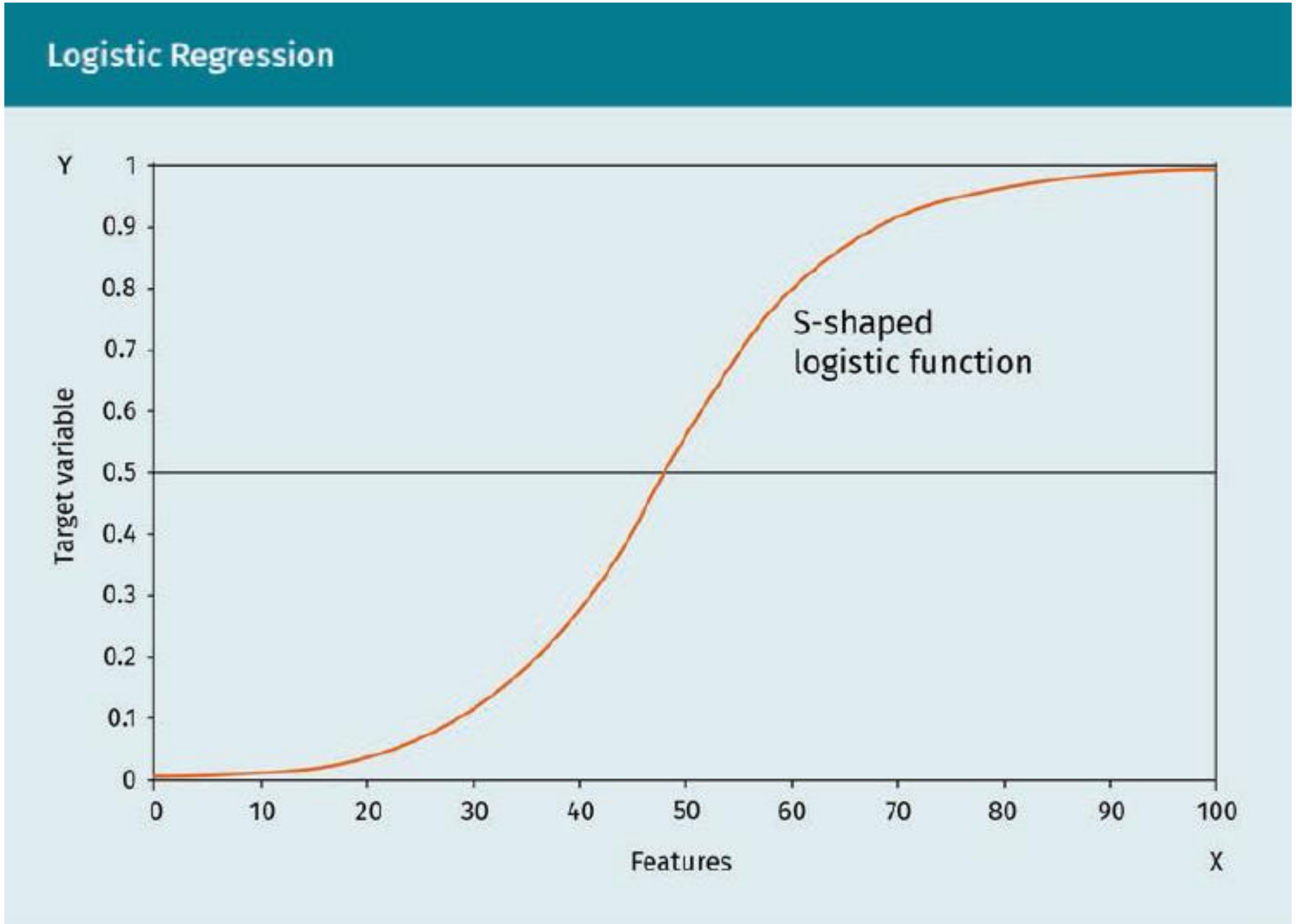


Source of image 4: Christian Müller-Kett, 2021.

REGRESSION

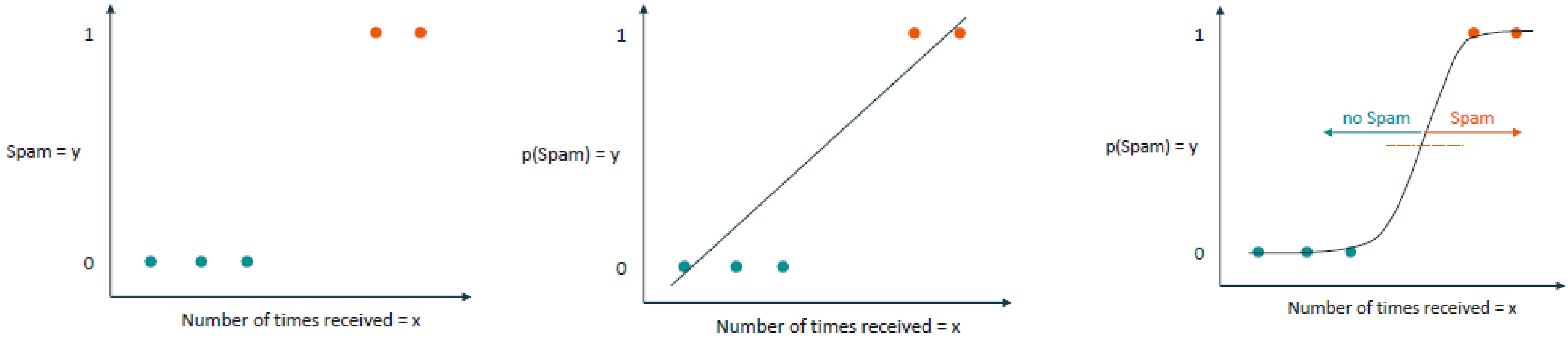

Linear Regression
Lasso and Ridge Regularization
Generalized Linear Models
Logistic Regression

LOGISTIC REGRESSION



Source of image: Course book

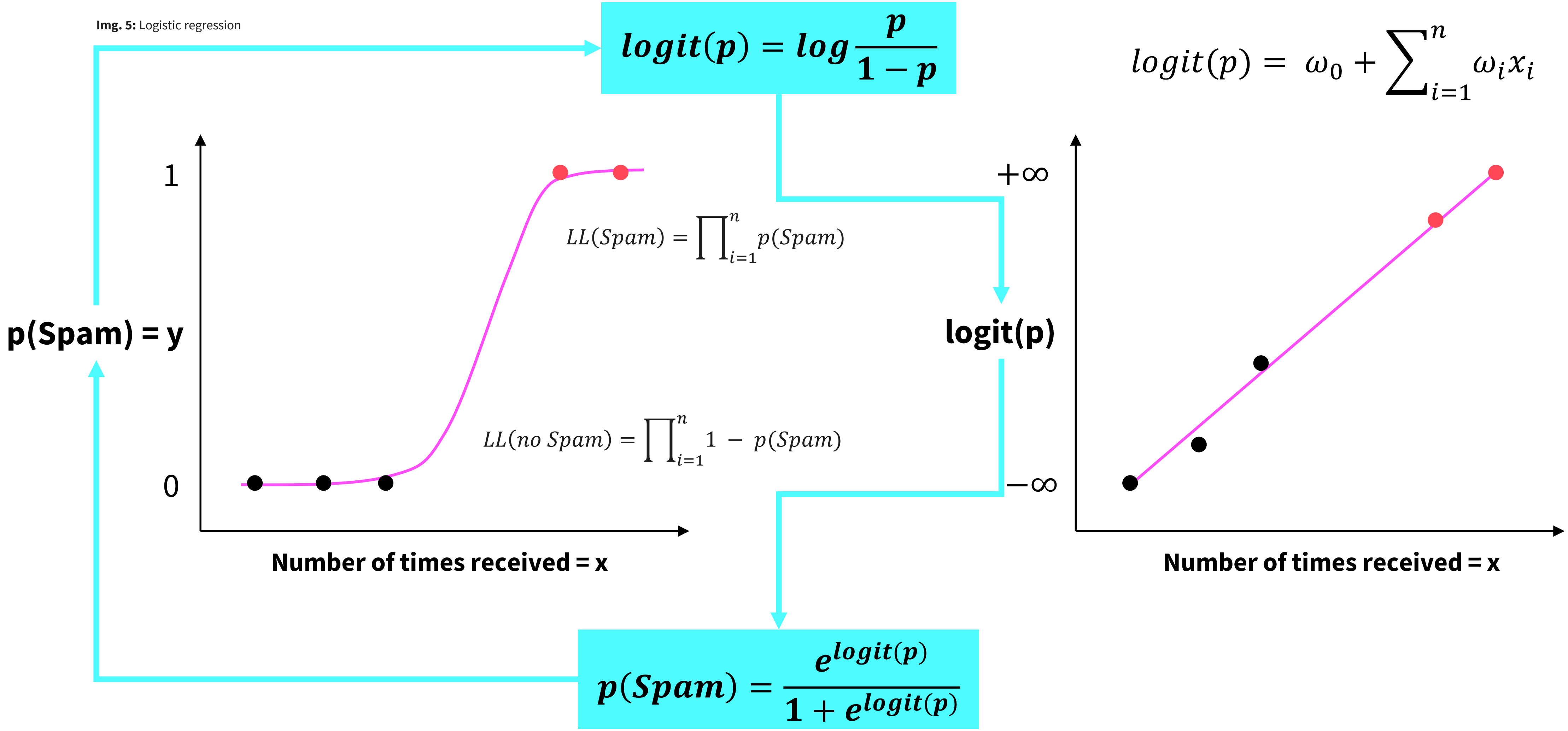
LOGISTIC REGRESSION



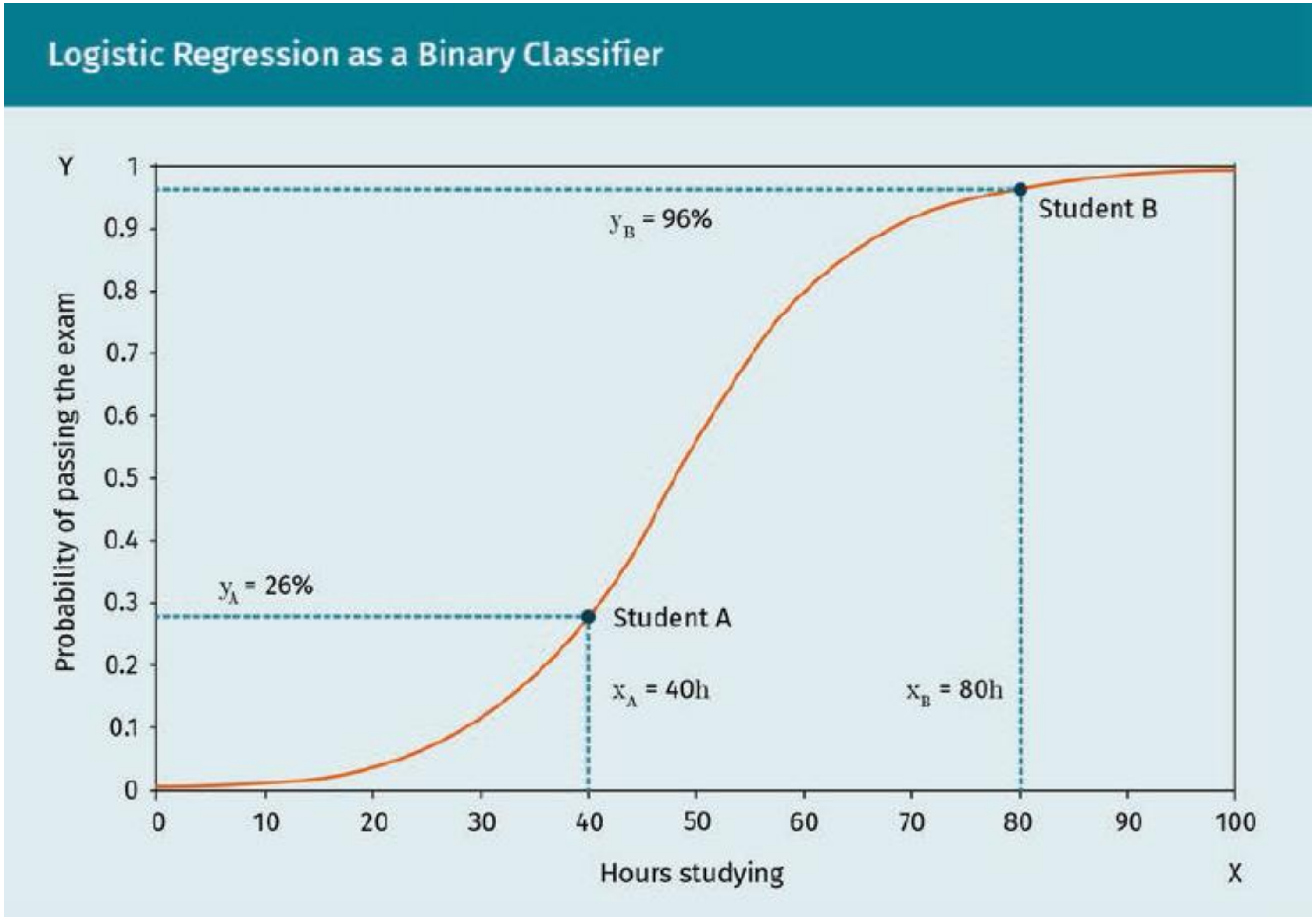
Source of image: https://learn.iu.org/

LOGISTIC REGRESSION

Img. 5: Logistic regression



Source of image 5: Christian Müller-Kett, 2021.



Logistic function

$$p\left(X\right) = \frac{e^{\omega_0 + \omega_1 X}}{1 + e^{\omega_0 + \omega_1 X}}$$

Multiple features

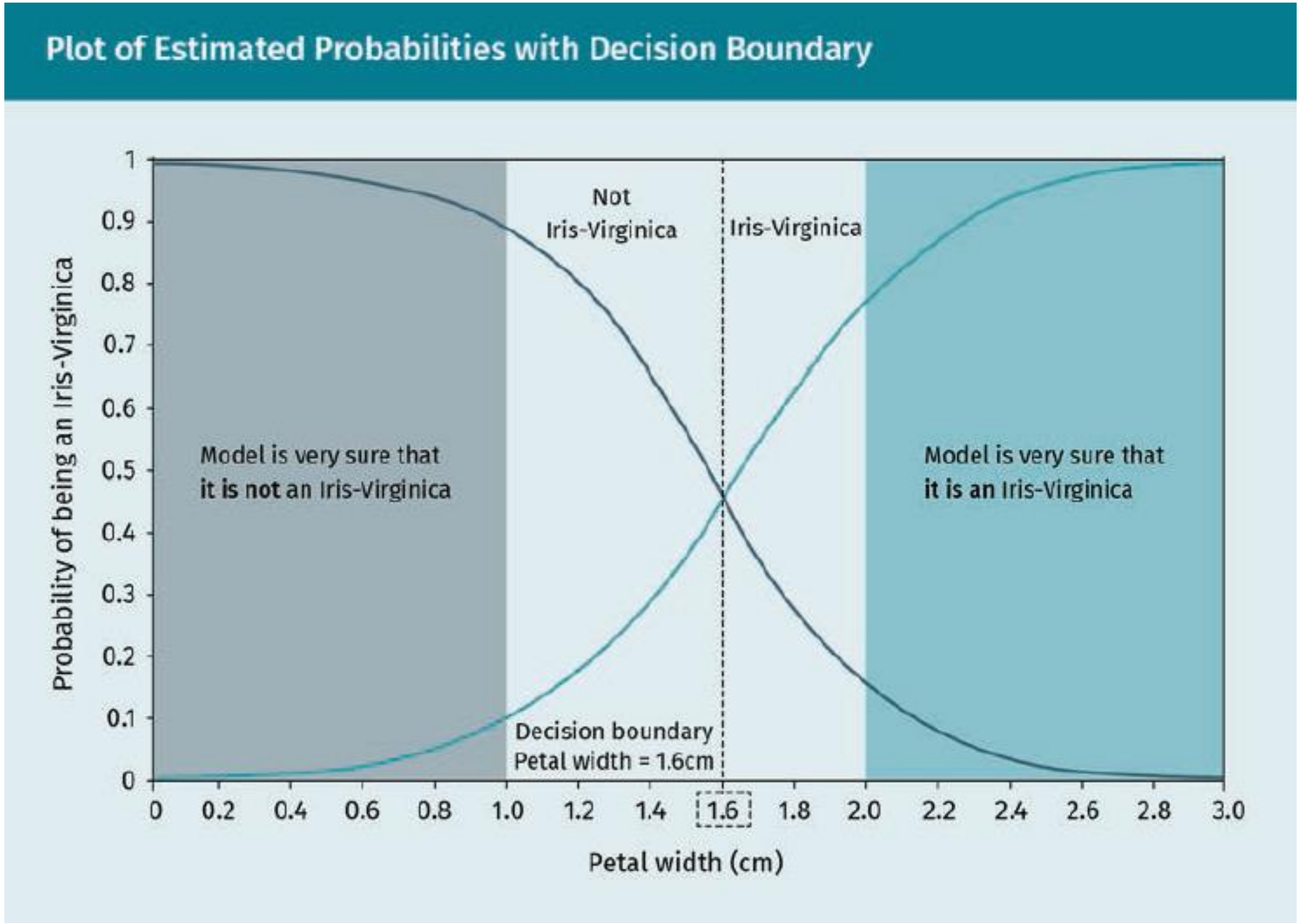
$$p\left(X\right) = \frac{e^{\omega_0 + \omega_1 X_1 + \ldots + \omega_p X_p}}{1 + e^{\omega_0 + \omega_1 X_1 + \ldots + \omega_p X_p}}$$

Source of image: Course book

- Model Training via Maximum Likelihood
  - Estimates for our regression coefficients **ω** such that the predicted probability **p(X)** for each observation matches the observed output value Y of the observation as closely as possible

$$l(\omega_0, \omega_1) = \prod_{i:\ y_i = 1} p(X_i) \prod_{i':\ y'_i = 0} [1 - p(x'_i)]$$

LOGISTIC REGRESSION – EXAMPLE



Source of image: Course book

REVIEW STUDY GOALS

- Understand the **concept** of regression and when to use it.
- Evaluate a regression model’s **performance**.
- Utilize **regularization** techniques and understand where they are implemented.
- Apply different well-known regression models with the use of **Python**.

SESSION 2

TRANSFER TASK

A start-up that sells **sustainable products in smaller stores** has been very successful in recent years. As a result, more stores are to be opened worldwide.

As a Data Scientist, you and your team are tasked with training a **machine learning model predicting product demand** one week ahead.

After fitting a first linear regression model, you realize that the **relationship** between features and the to-be-predicted demand seems **not** to be **linear**.

Also, the **training R^2** is **much higher** than the **testing R^2**.

Explain why this might be and **propose solutions** for these issues.

Please present your results.
The results will be discussed in plenary.

1. What is not a synonym for the loss function in a linear regression model?
   a) target function
   b) kernel function
   c) error function
   d) objective function

2. What is the name given to the deviations in a regression model?
   a) residuals
   b) support vectors
   c) eigenvalues
   d) estimates

3. Which one of the following is not a link function used in generalized linear models?
   a) identity
   b) index
   c) logit
   d) log