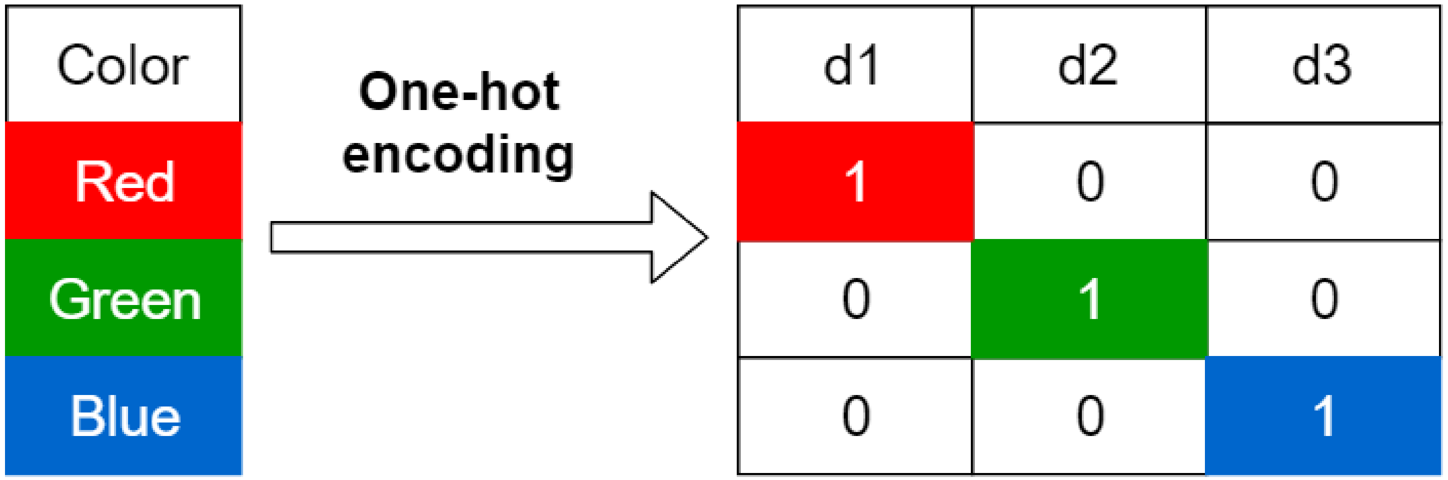
Dis5: hw2

Ziwen Kan
Slide adapted from Pu Sun

Announcement

• Group Formation	Jan 23 Thur	11:30PM
• Activity 1:	Jan 24 Fri	11:30PM
• HW2:	Jan 31 Fri	11:30PM

One-hot encoding



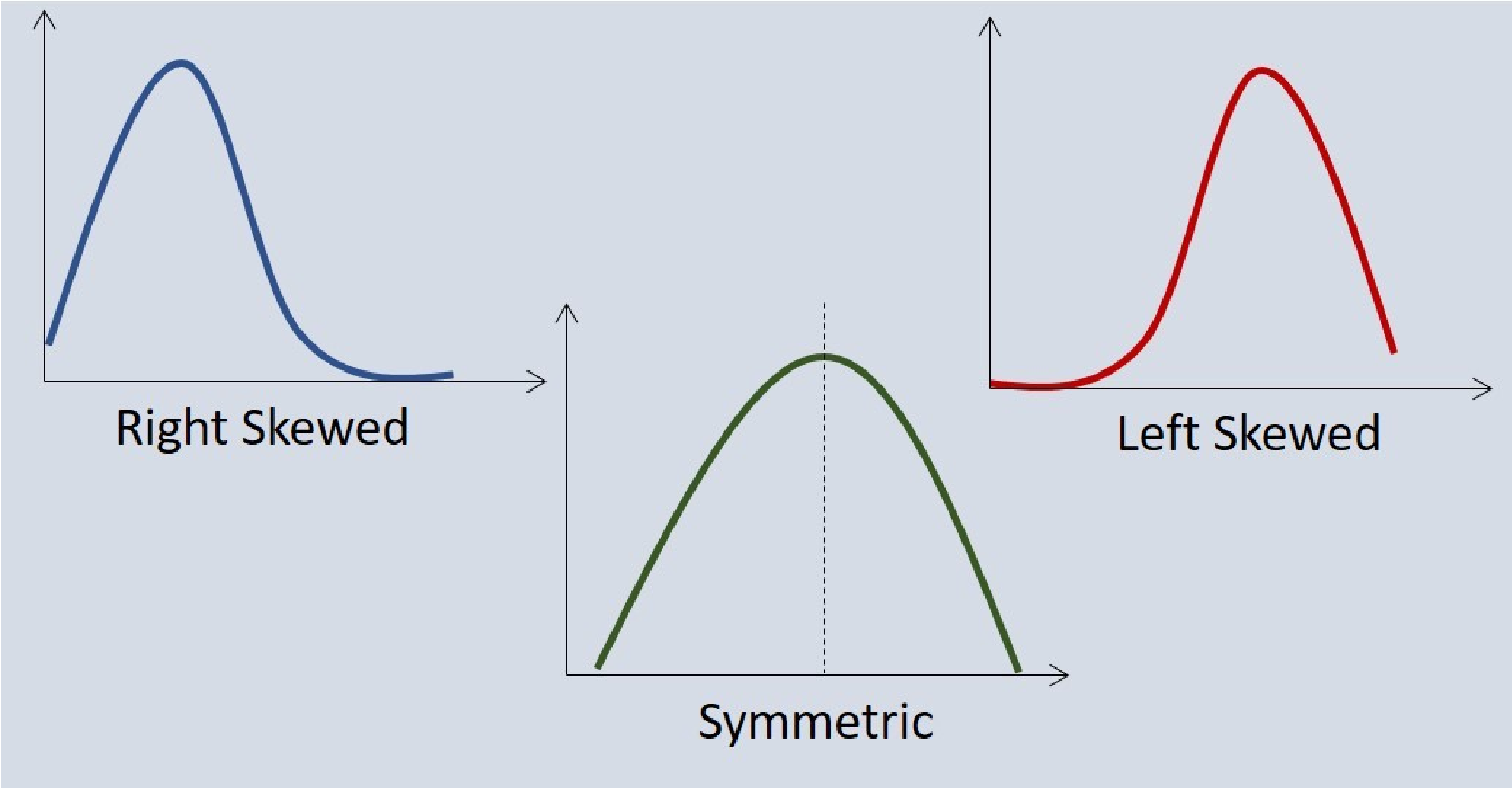
Min-Max scaler

- Normalization of Data
- Standardization of Features
- Preservation of Relationships
- Ease of Interpretability
- Improved Convergence Speed

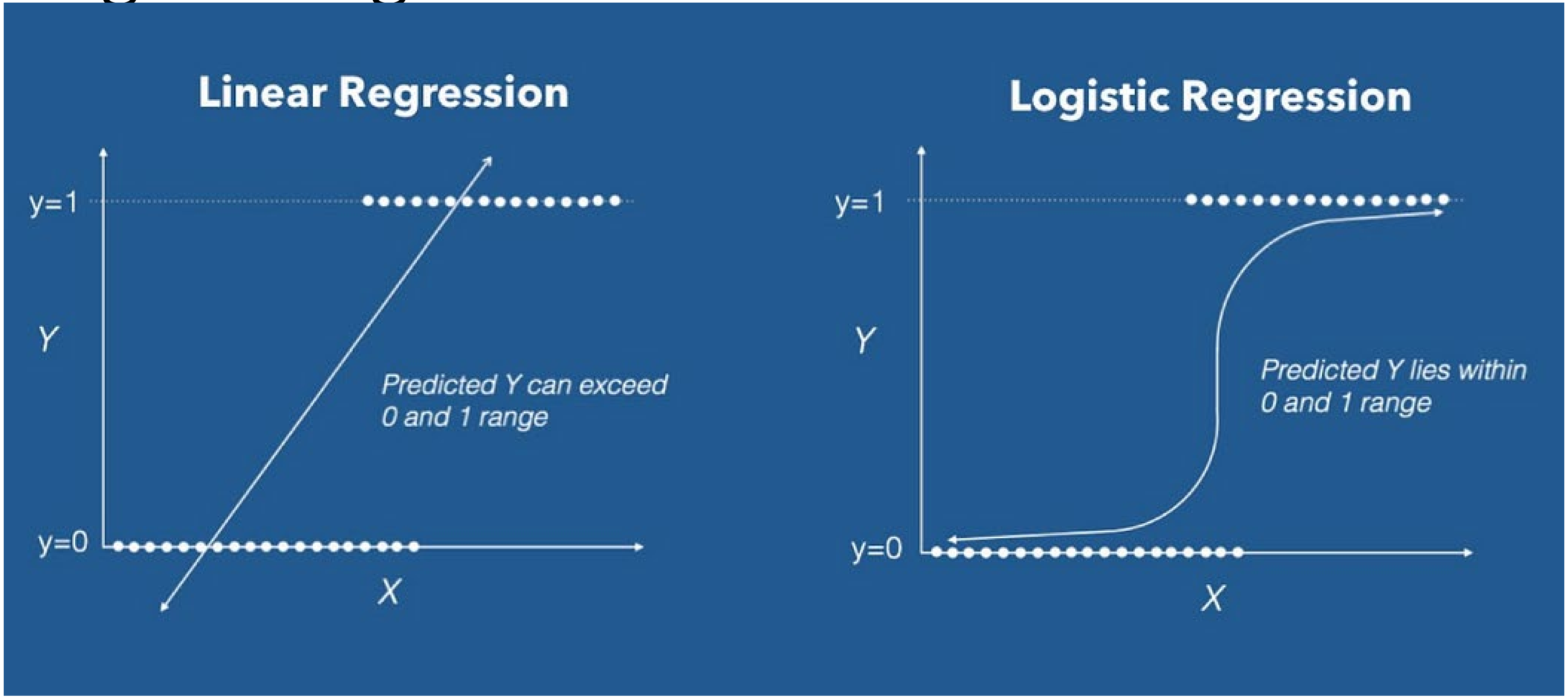
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Distribution

- Bimodal or multimodal if it is not able to be determine



Logistic regression



$$f(x) = \frac{1}{1 + e^{-x}}$$

RMSE and R^2

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

Measures the proportion of variance in the dependent variable that is explained by the independent variable(s) in a regression model.

Where,

$\hat{y}$ – *predicted value of y*

$\bar{y}$ – *mean value of y*

Log likelyhood $\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$

- What is it?

logarithm of the likelihood function, a fundamental concept in statistics that quantifies how well a statistical model explains a given dataset.

High Log-Likelihood: Indicates the model explains the data well (the data is probable under the model).

Low Log-Likelihood: Indicates the model poorly explains the data (the data is improbable under the model)

- Why?

Data points are assumed to be independent, their joint likelihood

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta)$$

For a normal distribution: a single data point's likelihood

$$f(y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

Why

Improves Interpretability

•**Transforms Tiny Values**: Converts small likelihood values (e.g., 10^{-50}) into manageable log-scale values (e.g., −50).

•**Easier Comparisons**: Simplifies evaluating model fit across datasets or parameter settings.

•**Highlights Contributions**: Summing log-likelihoods allows identifying individual data points' influence on model fit.

•**Readable Results**: Log-likelihood values are more human-readable and interpretable than raw likelihoods.

Example

Likelihood vs. Log-Likelihood

Consider three data points with probabilities 0.2, 0.1, and 0.05. The likelihood is:

$$L = 0.2 \cdot 0.1 \cdot 0.05 = 0.001$$

The log-likelihood is:

$$\log L = \log(0.2) + \log(0.1) + \log(0.05) \approx -0.698 - 1 - 1.301 = -2.999$$

The log-likelihood avoids dealing with the tiny value 0.001 directly and is easier to compute and interpret.

Log-likelihood in implementation

Take 10 as the base of the log.

Assuming \sigma is 1