DS

boosting your data exploration

Machine Learning Hands-on

KhanhPhamDinh, VinAI Applied, Ha Noi

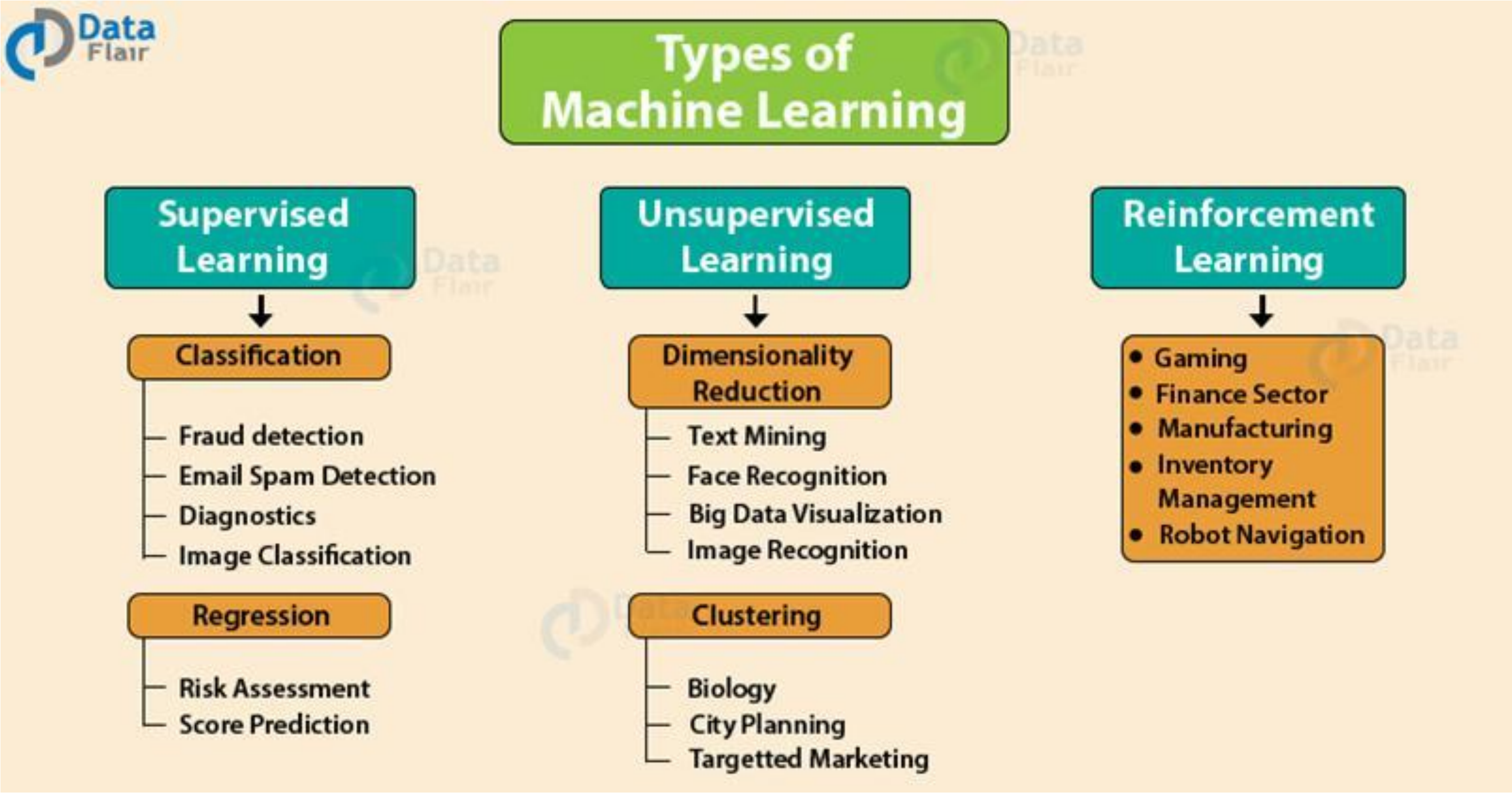
my book: Machine Learning Algorithms to Practices
https://phamdinhkhanh.github.io/deepai-book/intro.html

Logistic Regression

Content

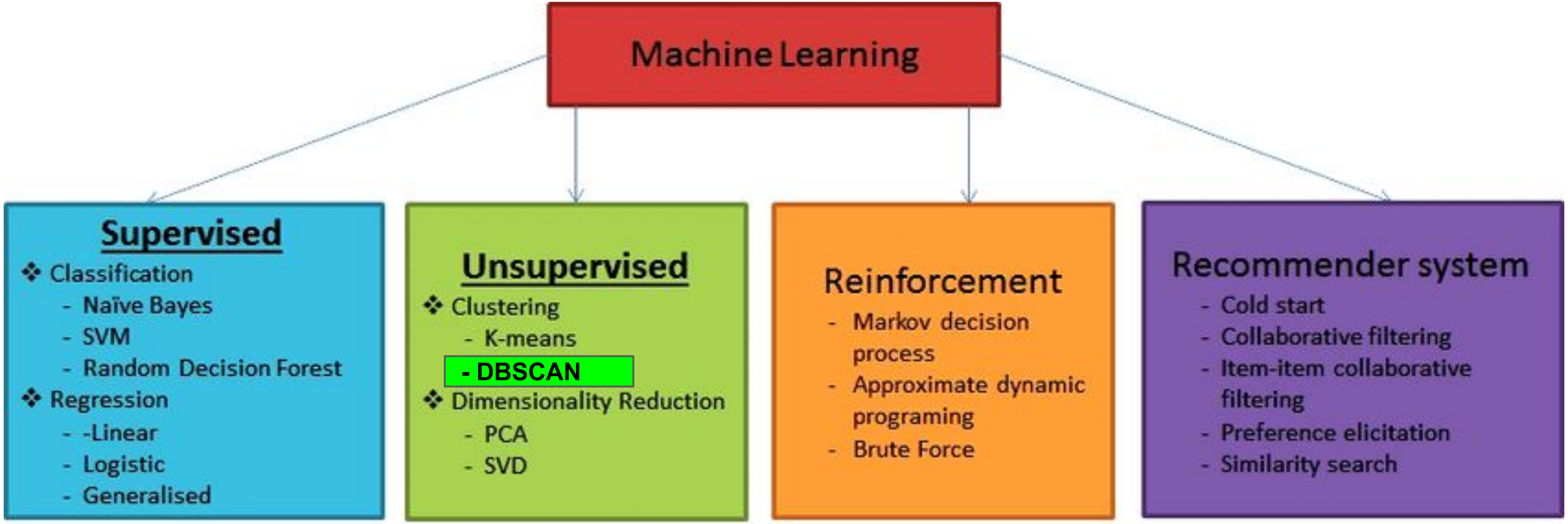
- General Machine Learning types
- Case Studies of Logistic Regression
- Why Logistic Regression?
- Sigmoid Function
- Cross Entropy - Loss Function
- Softmax regression

General machine learning types



https://data-flair.training/blogs/types-of-machine-learning-algorithms/

General machine learning types



Case Study of logistic regression

- Loan classifier:
  - Fraudulent vs Non-Fraudulent loans

- Breast cancers:
  - Cancer/not cancer

- Spam email:
  - Spam/Not Spam

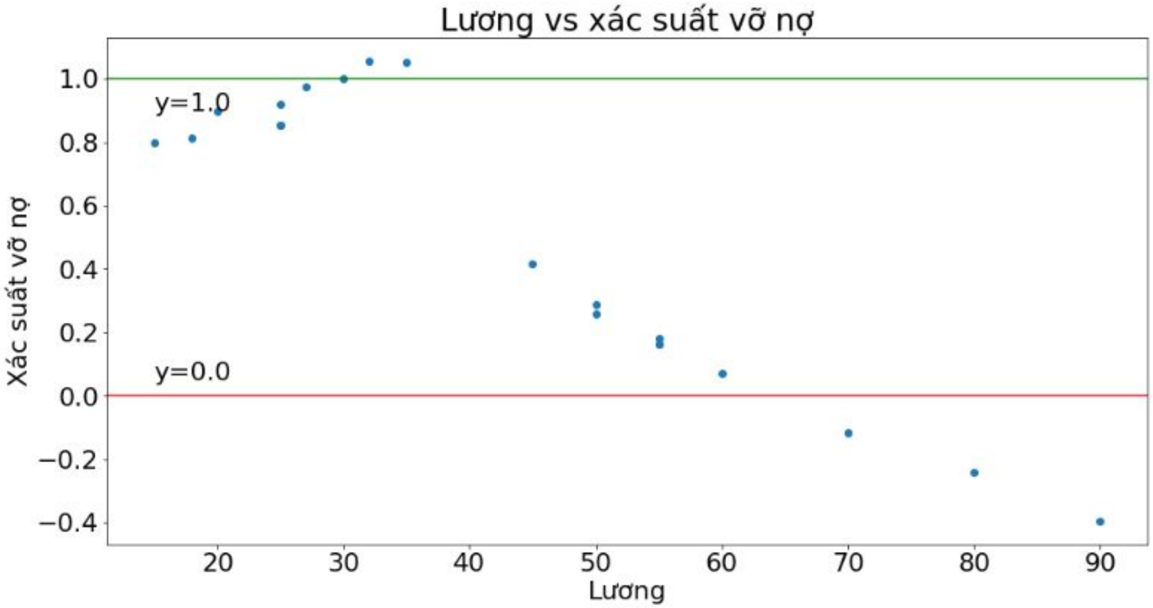
Why Logistic Regression

why does not use linear regression in classification?

Why Logistic Regression

why not using linear regression in classification?

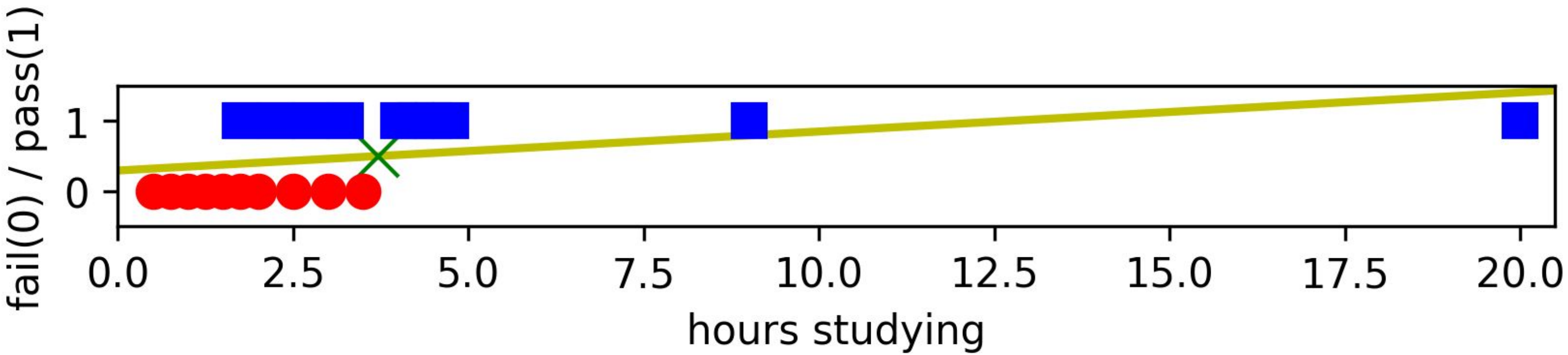
- Predicted values can be fall outside [0, 1].



Why Logistic Regression

why not using linear regression in classification?

- Sensitive with outliers.



- ❖ input: number of hours studying.
- ❖ target: pass exam ? pass = 1, fail = 0.
- ❖ threshold: y > 0.5 → pass, y <= 0.5 → fail.

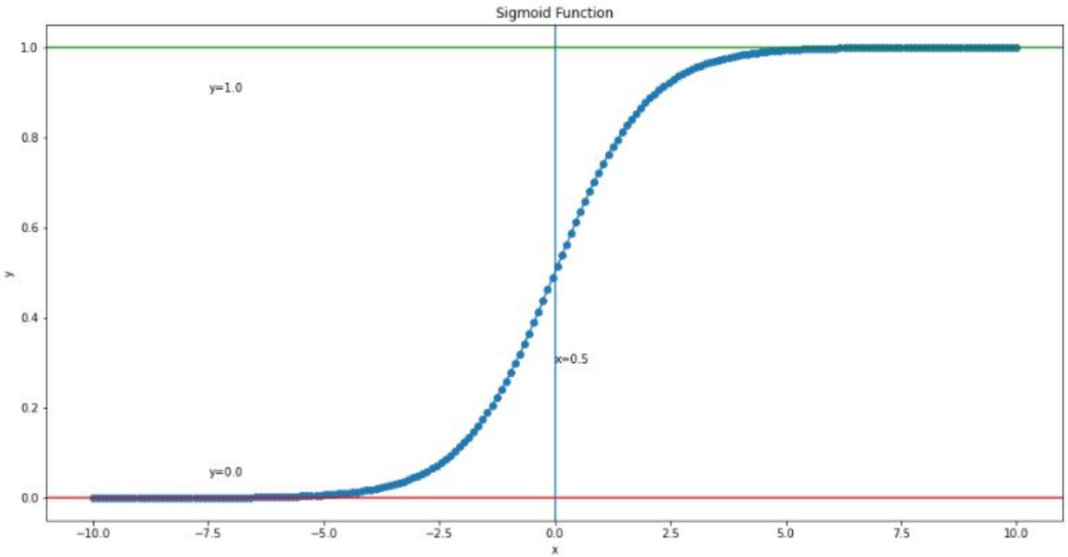
Why Logistic Regression

Sigmoid function

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

$$\lim_{x\to+\infty} \sigma(x) = \lim_{x\to+\infty} \frac{1}{1+e^{-x}} = 1$$

$$\lim_{x\to-\infty} \sigma(x) = \lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0$$



- ❖ Sigmoid values: between [0, 1].
- ❖ It is non-linear regression.

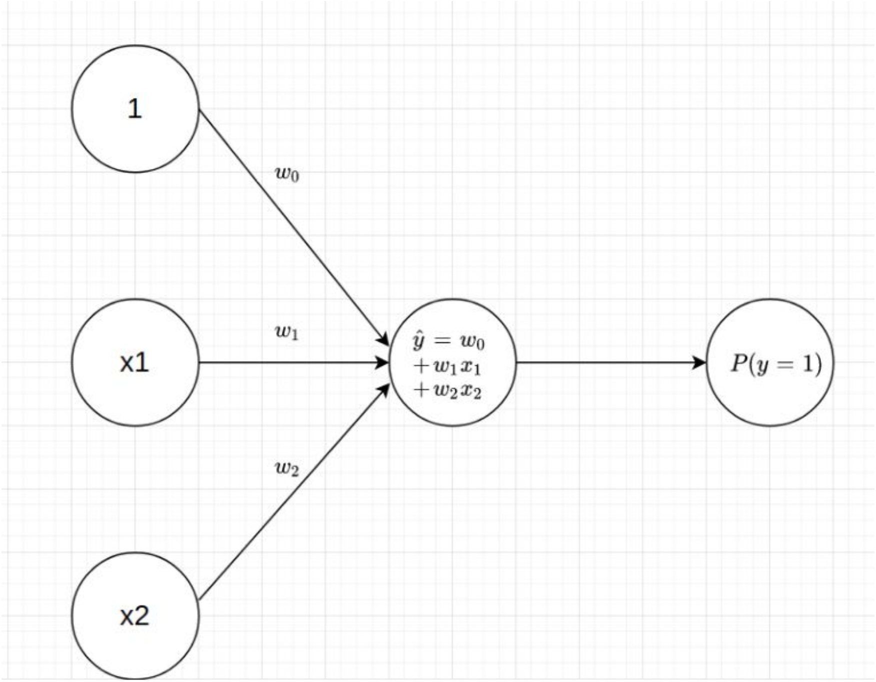
Why Logistic Regression

- Sigmoid function is simple.
- It has a high explainability compare with other algorithms.
- We can explain prediction probability based on the input variable.
- That is why is prefered in several business models such as Fraudulent detection, Fraudulent transaction,....

Logistic Regression

Graphical model

- x1, x2 are input variables and 1 is additional value.
- Logistic Regression Graph:
    - step 1: linear combination.
    - step 2: forward to sigmoid non-linear function
- Output:
  forecast probability P(y=1 | **x**)

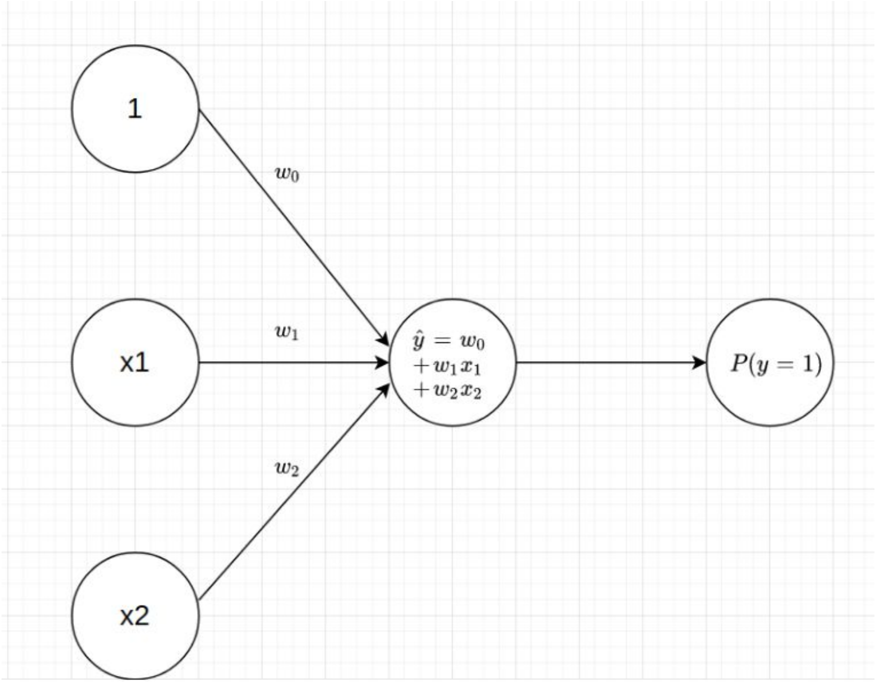


Logistic Regression

Setup Probability threshold for output label

- 0.5 is default threshold

$$\begin{cases} 0 \text{ if } P(y=1|\mathbf{x};\mathbf{w}) \leq 0.5 \\ 1 \text{ if } P(y=1|\mathbf{x};\mathbf{w}) > 0.5 \end{cases}$$



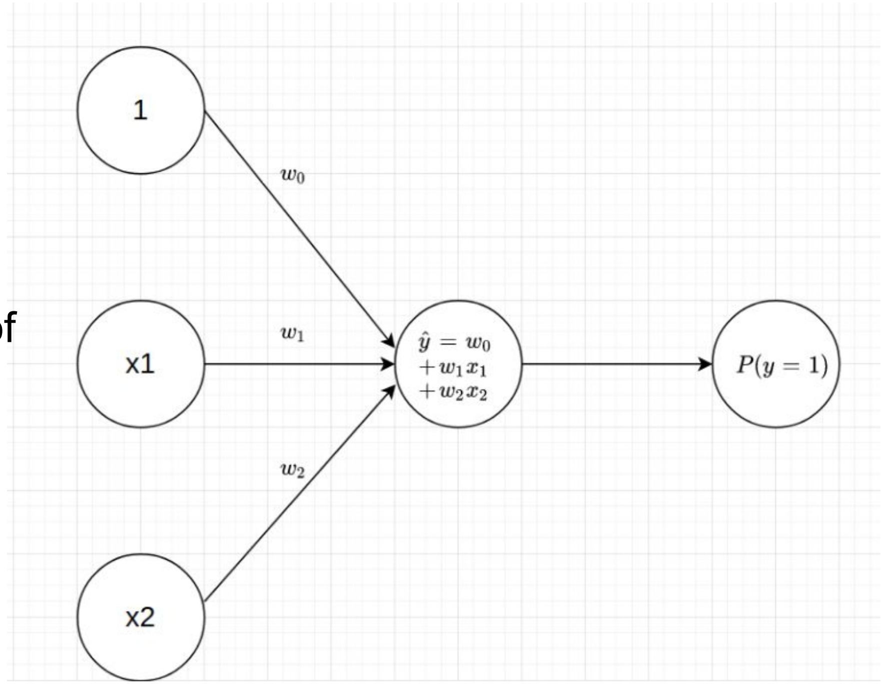
Logistic Regression

Setup Probability threshold for output label

- ❖ 0.5 is default threshold

$$\begin{cases} 0 \text{ if } P(y=1|\mathbf{x};\mathbf{w}) \leq 0.5 \\ 1 \text{ if } P(y=1|\mathbf{x};\mathbf{w}) > 0.5 \end{cases}$$

- ❖ threshold can be change → outputs of model going to be change.



Logistic Regression

Decision boundary

❖ h_w(x) is a hypothesis function that is used to predict P(y=1).

label 0

$$
\begin{aligned}
h_{\mathbf{w}}(\mathbf{x}) &\leq 0.5 \\
&\leftrightarrow \frac{1}{1+e^{-\mathbf{w}^\top \mathbf{x}}} \leq 0.5 \\
&\leftrightarrow e^{-\mathbf{w}^\top \mathbf{x}} \geq 1 \\
&\leftrightarrow \mathbf{w}^\top \mathbf{x} \leq 0
\end{aligned}
$$

label 1

$$
\begin{aligned}
h_{\mathbf{w}}(\mathbf{x}) &> 0.5 \\
&\leftrightarrow \frac{1}{1+e^{-\mathbf{w}^\top \mathbf{x}}} > 0.5 \\
&\leftrightarrow e^{-\mathbf{w}^\top \mathbf{x}} < 1 \\
&\leftrightarrow \mathbf{w}^\top \mathbf{x} > 0
\end{aligned}
$$

Logistic Regression

Odd ratio

❖ odd ratio is a proportion between positive and negative probability.

$$\text{Odd Ratio} = \frac{P(y=1|\mathbf{x};\mathbf{w})}{P(y=0|\mathbf{x};\mathbf{w})} = \frac{P(y=1|\mathbf{x};\mathbf{w})}{1-P(y=1|\mathbf{x};\mathbf{w})} = e^{-\mathbf{w}^\intercal\mathbf{x}}$$

Logistic Regression

Odd ratio

- odd ratio is a proportion between positive and negative probability.

$$\text{Odd Ratio} = \frac{P(y=1|\mathbf{x};\mathbf{w})}{P(y=0|\mathbf{x};\mathbf{w})} = \frac{P(y=1|\mathbf{x};\mathbf{w})}{1-P(y=1|\mathbf{x};\mathbf{w})} = e^{-\mathbf{w}^{\intercal}\mathbf{x}}$$

Logistic Regression

Loss function

[*7.Logistic Regression - Lecture Note*]

Softmax Regression

Softmax function

- ❖ applied for multiple classification.

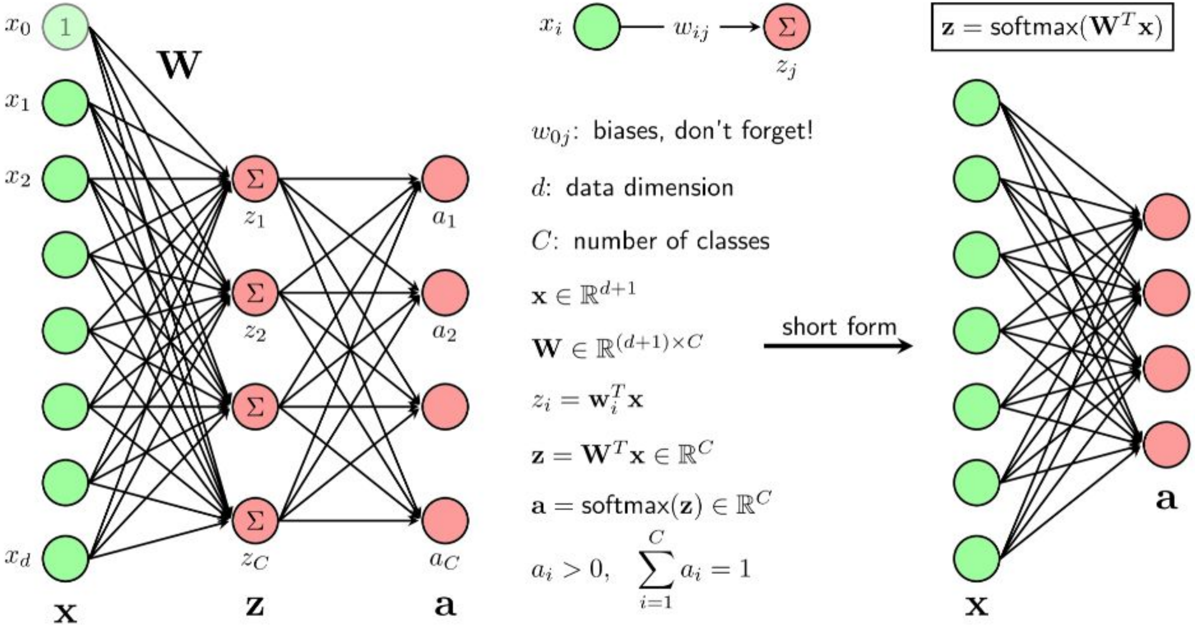
$$a_i = \frac{\exp(z_i)}{\sum_{j=1}^{C} \exp(z_j)}, \quad \forall i = 1, 2, \dots, C$$

- ❖ Estimate probability for each class:

$$P(y_k = i | \mathbf{x}_k; \mathbf{W}) = a_i$$

Softmax Regression

Softmax function



Hình 2: Mô hình Softmax Regression dưới dạng Neural network.

https://machinelearningcoban.com/2017/02/17/softmax/

Softmax Regression

Softmax function - avoid overflow

- ❖ c can be minimum values of **z.**

$$\frac{\exp(z_i)}{\sum_{j=1}^{C}\exp(z_j)} = \frac{\exp(-c)\exp(z_i)}{\exp(-c)\sum_{j=1}^{C}\exp(z_j)} = \frac{\exp(z_i - c)}{\sum_{j=1}^{C}\exp(z_j - c)}$$

References

1.Logistic Regression - khanhblog

2. Logistic Regression - Machine Learning Cơ bản

3. Standford Logistic Regression

4. Standford Lecture Note: 23-LogisticRegression.pdf

5. Bishop Pattern Recognition - Logistic Regression