Descriptive power

Zhihan Hu

February 2025

In machine learning, descriptive power is defined as the ability of a model to capture complex relations (patterns) between input features and outputs. Though an abstract concept that lacks theoretical generalization, descriptive power of a model can primarily improved by schemes presented later in the document. I will also provide intuitive reasons why each scheme has the ability to improve descriptive power.

1 Adding parameters

In machine learning, learnable parameters represent the patterns that are going to be captured. By increase the number of learnable parameters, the dimension of the pattern space will increase, giving more freedoms and capacity to learn the patterns.

Examples In linear regression, we basically have a row vector linear map for each feature, i.e

$$y = \theta^T x + b \tag{1}$$

where

- x $\in R^{n\times 1}$
- $\theta \in R^{n\times 1}$

If we are going to add another weight matrix W $inR^{n\times n}$, the new linear regression will be

$$y = \theta^T W x + b \tag{2}$$

By adding n^2 more parameters to the model, the degrees of freedom and capacity will be improved, and thus may improve the potential to capture a more accurate patterns.

2 Adding complexities

2.1 Add non-linearity

Adding parameters may help to improve descriptive power; however, the influence might not be outstanding. Based on the example provided above, even we keep adding more weight matrices W_1, W_2..., the model will remain linear, meaning that if there exists non-linear relations in between features, this model will never capture. Therefore, for current ml models, non-linearity is often added to the model to capture non-linear relations.

Examples:
In modern deep learning model, there will be an activation process such as ReLU, sigmoid, softmax, tanh for handling the output from one layer. Based on universal approximation theorem, these non-linear function can help to approximate any continuous functions.

2.2 Add layers

By adding layers, an increased number of learnable parameters and more complicated non-linear relations can be represented by a model, leading to better descriptive power.

2.3 Independent capturing of different patterns

This is a technique that is designed to capture different patterns independently instead of learning all different patterns within same weight matrix. The primary architect of such technique is

- initial multiple sets of weight matrices
- apply each matrix independently and in parallel on the input
- Use a redundancy matrix O to integrate learned patterns from each weight matrix
- adding non-linearity in the intermediate layers

The name of this technique is called multi-view representation learning. When I first learned about this concept, I was confused that this architect guarantees that each weight matrix can capture different patterns? The answer is, actually there is no guarantee, but this architect increases the descriptive power by a more complex relations and thus has a potential to capture more complex patterns.

Example:
Multi-head self-attention

3 Process of features

Not only model itself, the feature itself can contribute to improve the descriptive power. For example, if we add higher order or interactive terms, we may be able to find potential patterns that is related to these terms.

Example:

- Support vector machine
- A much higher neuron vector dimensions among intermediate layers in a deep learning model.