Unified framework in machine learning
A report by

Nìm Trí Nghĩa

July 2019

1 General framework

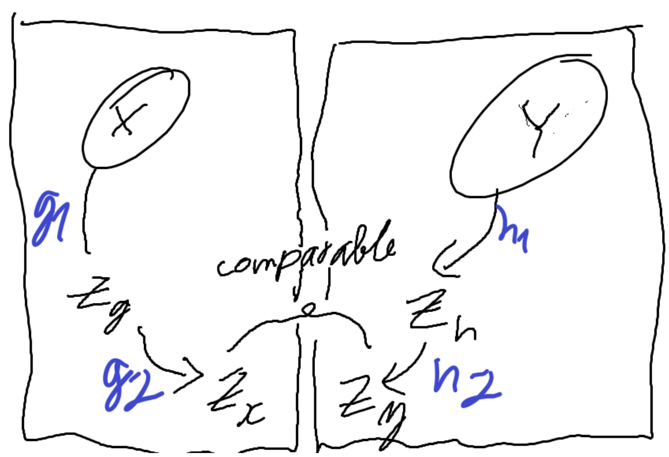


Figure 1: **Sketch of a unified framework in machine learning**

How it works: Unlike human, a machine does not possess the ability to differentiate between things in a high-dimension space. In order for it to rate its performance, the information must be strip down to a lower dimension space, called "**embeddings**".

The 'X' and 'Y' spaces represents the information that we want the machine to choose and extract; we can think of this as a **vector space**. Each vector is an attribute, called a **basis feature**, that a computer can extract from. Since the amount of vectors in the space can be enormous, and the capacity of a machine is by contrast limited, it must only choose some of the vectors available that we need.1

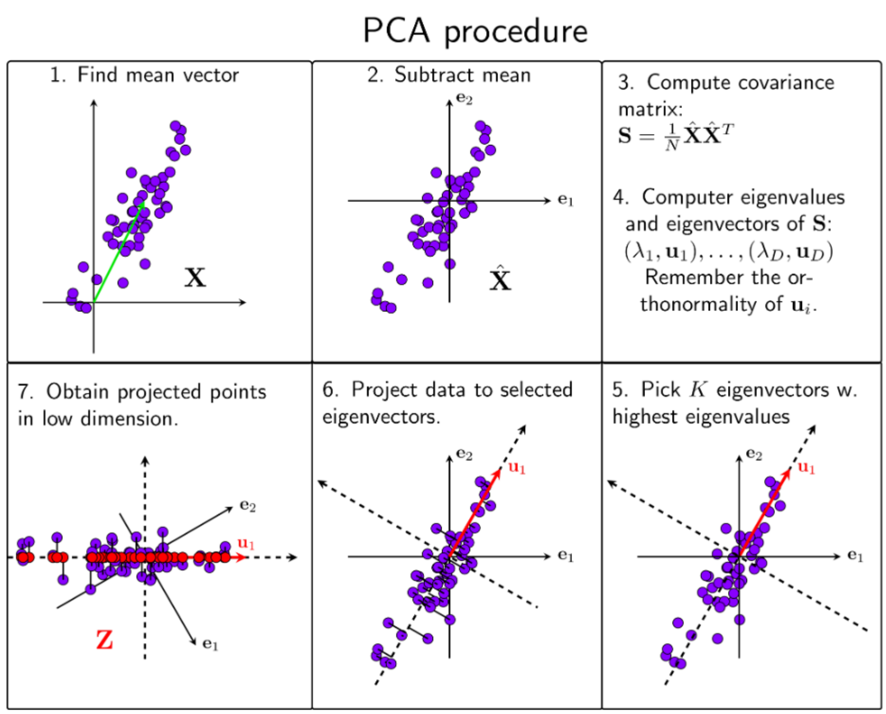
For example: we want the a machine to predict the taste of an ice-cream. To do so, we need it to pick out the color, the smell,... we don't need it to figure out what type of holder is holding the ice-cream, nor what type of person is holding the ice-cream.

In the above example, the X space would be the information about the ice-cream and the Y space would include words from a dictionary.

The way we can extract the necessary information (color, smell,...) and words (sweet, bitter, strawberry, crunchy...) is by using some **basis function**. Each basis function corresponds to a specific type of information, e.g., color. g_1, g_2, h_1, h_2 in the sketch are a collection of these basis functions. These produce a **coordinate vector Z** containing the picked-out information in both the X and Y vector space.

2 Principal Component Analysis

This is the simplest way we can reduce our high-dimensional vector space to a lower one. From our initial vector space, we can use this method to reduce it to the coordinate vector Z_g and Z_n in our sketch



After this step, if data from Z_g and Z_n can be expressed in the same vector space, we can compare their performance through either:

- Calculating the Inner product
- Calculating its distance between reality and machine-produced (e.g., real flavour to the flavour that the code gives).1
- Many other methods depending on the type of data like Cross-entropy...

Else, we move on to the next step, which is to strip it down to an even lower dimension space.

3 Linear regression

This is a method used for specific types of information in statistics.

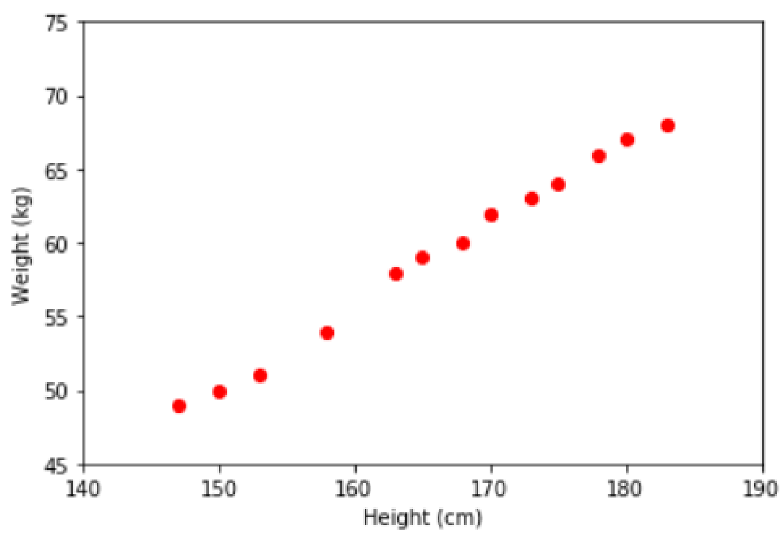


Figure 2: **Dataset of height and weight of 13 people**

This is one of the method that can be used to find the optimal functions for the next step after **PCA**, after which, it is stripped down to its simplest components and can be compared

As the name suggests, it is used mainly for regression problems

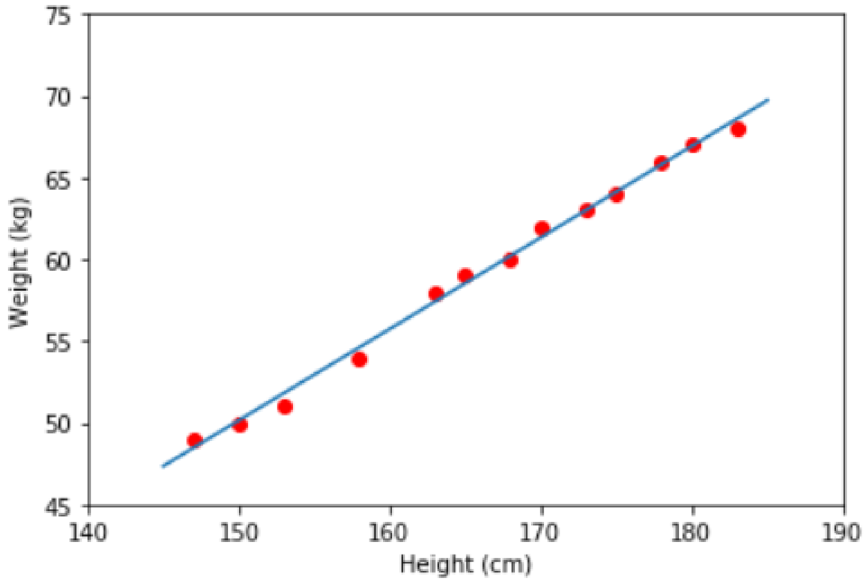


Figure 3: **Using linear regression to compute the line best fit**

4 Logistic regression

Despite its name, it's not used mainly for regression problems, but instead, it is more compatible with classification problems. This is because the output of logistic regression is a probability, making it suitable to identify things.