Introduction to Modeling

H.G Min

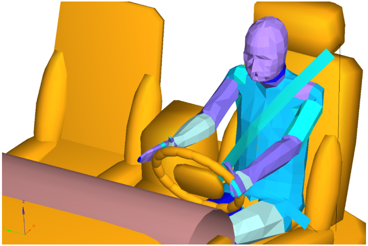
Yonsei University, Korea Quantum Computing

mhg9511@gmail.com

August 14, 2022

Overview

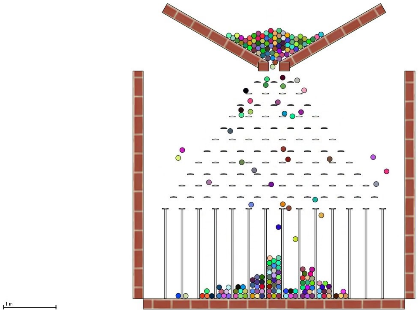
What is Modeling?



Modeling is "**making theoretical structure to explaining the real data**"
We can use the model to 1.**Analyze**, 2.**Predict** and 3.**Generate** the data.

Most of the modeling have following components

- D : Data
- $f(\cdot;\theta)$: Model Structure
- θ : Parameter
- $c(f, D, \theta)$: Criterion to measure the performances of the models



Predictive Model

Most of the machine learning models are target to predict response Y when it gives predictor X

$$Y \sim X$$

Then, the components become as follow.

$$\begin{aligned} D \Rightarrow & \quad (X_i, y_i)_{i=1}^n \\ f(\cdot;\theta) \Rightarrow & \quad f(X_i;\theta) \\ c(f, D, \theta) \Rightarrow & \quad d(y_i, f(X_i;\theta)) \end{aligned}$$

Predictive Model

Most of the machine learning models are target to predict response Y when it gives predictor X

$$Y \sim X$$

Then, the components become as follow.

$$\begin{aligned} D \Rightarrow & \qquad (X_i, y_i)_{i=1}^{n} \\ f(\cdot;\theta) \Rightarrow & \qquad f(X_i;\theta) \\ c(f, D, \theta) \Rightarrow & \qquad d(y_i, f(X_i;\theta)) \end{aligned}$$

Minimal Mean Squared Error

The most conventional choice of the cost function is **Mean Squared Error (MSE)**.

$$E(\mid y_i - f(X_i;\theta) \mid^2)$$

We can see that the MSE become minimum when $f(X_i;\theta) = E(y \mid X_i)$.

Therefore, most of the predictive models are target to build and find the **function** f to estimate $E(y \mid X_i)$.

$\bigoplus$ Notice that this is only a point estimation. Many statistician are also interested in finding more general inferences of the Y such as **Interval Estimation** and **Distribution Estimation**.

Functional Space

At the most fundamental phase of the modeling, there is a concept called as functional space. A function is called as **functional** if the codomain of the function is a scalar field such as $\mathbb{R}$ or $\mathbb{C}$.

A functional space could be understood as a set of all functional, i.e.
$F = \{f; f : M \to \mathbb{R}\}$

$$ex)x, 2x, x^2, \exp(x), \log(x), \cos(x), \sin(x) \in F$$

This is too flexible and ambiguous to handle!
Let's use more easy and tractable space $\Rightarrow$ **parameterizing techniques.**

Parameterized Techniques

Parameterized Techniques : use more easy components called parameters such as $\mathbb{R}^p$ to handle the functions.

Ex)

$$F^{(1)} = \{f; f(x;\beta) = \beta_0 + \beta_1 x, \beta_0, \beta_1 \in \mathbb{R}\}$$

Notice below properties.

1. Every element of $F^{(1)}$ is uniquely defined if (β_0, β_1) is fixed.
2. We can handle all of the functions in $F^{(1)}$ by handling the real vector (β_0, β_1)
3. $F^{(1)} \subset F$

We can build **linear model** using this space.

$$E(y \mid x) = f(x;\beta) = \beta_0 + \beta_1 x$$

Defining the base functional space = Defining the model structure

Ex)

$$F^{(2)} = \{f; f(x;\beta) = \beta_0 + \beta_1 x + \beta_2 x^2, \beta_0, \beta_1, \beta_2 \in \mathbb{R}\}$$

Notice below properties.

1. Every element of $F^{(2)}$ is uniquely defined if $(\beta_0, \beta_1, \beta_2)$ is fixed.
2. We can handle all of the functions in $F^{(1)}$ by handling the real vector $(\beta_0, \beta_1, \beta_2)$
3. $F^{(1)} \subset F^{(2)} \subset F$

We can always improve the expressibility of the model by adding nonlinear mapping term. $\Rightarrow$ **Higher Order Mapping Techniques**.

Let's see generalized version of the above example.
Ex)

$$F^{(P)} = \{f; f(x; \beta) = \sum_{k=0}^{p} \beta_k x^k, \beta_k \in \mathbb{R}\}$$

Notice below property.

- $F^{(1)} \subset F^{(2)} \subset ... \subset F^{(p-1)} \subset F^{(p)} \subset ... \subset F$

We can build the mathematical procedure to improve the expressibility of our models

Then, there is a question arising : Is $F^p \to F$ as $p \to \infty$?.

The Answer is No.

$F^p \to C^\infty$: Infinitely Differentiable Functional Space, Smooth Functional Space or Analytic Functional Space.

At least, we can know that our structure $\approx$ model $\approx$ functional space could express all of the functions in C^∞.

Figuring out some structures could express all of the element in the more general space is called problem of **universality**.

What is Good Model Structure?

Setting the model Structure = Setting the Functional Space

The criterion of Good Model Structure.

1. Whether it is **Universal** at the bigger space
2. How **Fast** it converges into bigger Space
3. How **Easy** it is to find the parameters

Kernel Space

We can generalize almost every machine learning model into linear combination of basis function.

$$f(x;\theta) = \sum_{i=1}^{p} \theta_i e_i(x)$$

Especially, the space with below structure called as kernel space.

$$f(x;\theta) = \sum_{i=1}^{n} \theta_i k(x, x_i)$$

Ex)

Neural Network Space

$$f(x; w) = \sum_{i=1}^{p} w_i^{(2)} \sigma_i(w_{ij}^{(1)} x)$$

Gaussian Kernel Space

$$f(x; \beta) = \sum_{i=1}^{n} \beta_i \exp(| x - x_i |^2)$$

Quantum Kernel Space

$$f(x; \beta) = \sum_{i=1}^{n} \beta_i \mid \langle \psi(x) | \psi(x_i) \rangle \mid^2$$

Application Examples

Kernel Regression

$$f(x) = \frac{\sum_{i=1}^{n} k(x, x_i) y_i}{\sum_{i=1}^{n} k(x, x_i)} = \sum_{i=1}^{n} \frac{y_i}{\sum_{i=1}^{n} k(x, x_i)} k(x, x_i)$$

Kernel Support Vector Machine

$$f(x) = sign\left(\sum_{i \in N_S} \alpha_i y_i K(x, x_i) + b \right)$$

Kernel Principal Component Regression

$$f(x) = \sum_{i=1}^{n} \beta_i k(x, x_i)$$

where $\beta = \arg\max \beta^t K^2 \beta$ subject to $\beta^t K \beta = 1$