Basics

Honghu Xue

Independent and identically distributed random variables

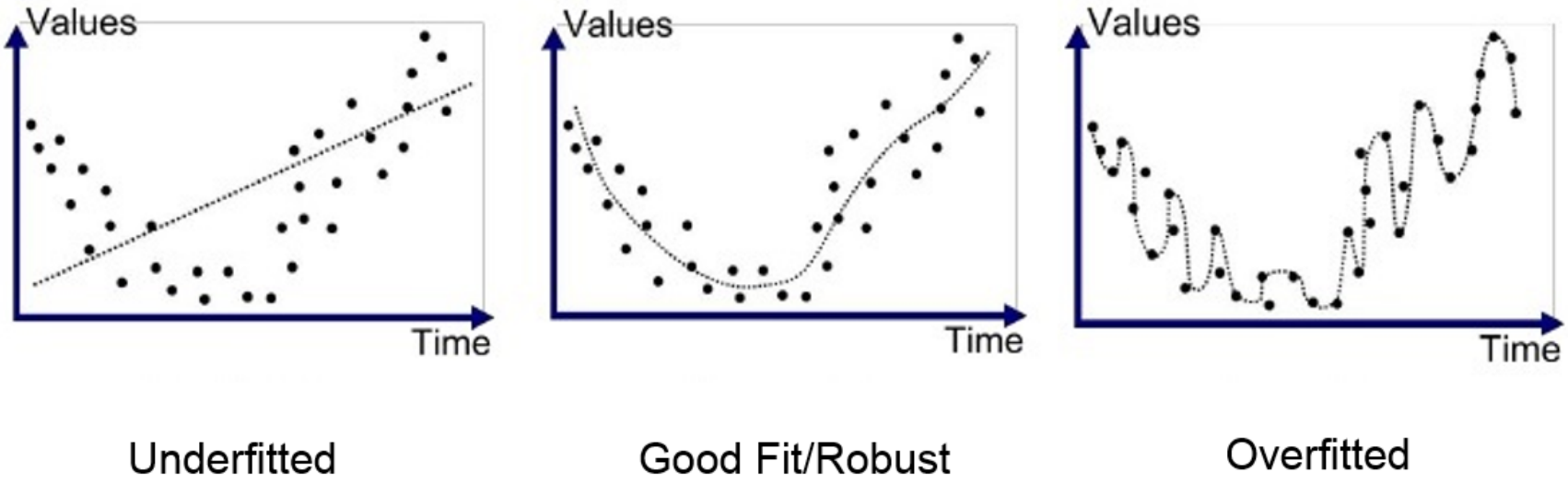
- a collection of random variables is **independent and identically distributed** if each random variable has the same probability distribution as the others and all are mutually independent.

- Examples of i.i.d. : Sample with replacement (the result of the n-th draw)

- Independent, non-identical distribution: the lottery changes its content after each draw

- Dependent, non-identically-distributed : temporal correlation, the sample from the previous time step affects the outcome of next sample.
- E.g. sample without replacement (take a draw from lottery, but not putting the draw back.)

Underfit v.s. Overfit



- The rule of Occam's Razor
  - Apply the simplest model which can explain the data → generalize well

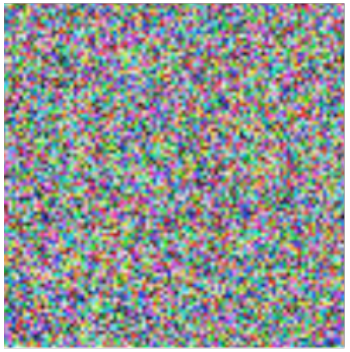
Generative model v.s. Discriminative model

- X : data instances, Y : Labels
- **Generative models** capture the joint probability p(X, Y), or just p(X) if there are no labels.
- E.g. Generative-Adverserial Networks , can generate new photos of animals that look like real animals

- **Discriminative models** discriminate between different kinds of data instances, ignores the question of whether a given instance is likely.
- Only captures the conditional probability p(Y | X)
- E.g. Logistic regression, decision tree , can only tell a dog from a cat

An illustration of GAN

Noise ~ N(0,1)



Generative Model

Data

- https://github.com/NRottmann/Toolbox-GP-GMRF