Wang Yanan Institute for Studies in Economics (WISE)

Discriminative & Generative
A Preliminary Study

Yufeng Gu & Huajin Yu

March 15, 2019

Abstract

Discriminative and generative are two fundamental concepts, which go through the whole process of machine learning. This report provides a preliminary approach to comparing these two things, including their backgrounds, principles, objectives, and probability insights. It's worth noting that when we expect two variables to correlate with each other, the generative models will always outperform discriminative models.

1 Introduction

As Devesh Batra (2016) noticed, economists use various classifiers that are suitable for operation on their dataset, so as to achieve one of the following objectives:

(i) To predict the targets based on the features of individuals;
(ii) To generate the causal effects of features on targets.

While there exist quite a lot of approaches, classifiers can be broadly divided into two groups according to the two objectives (Wikipedia, the free encyclopedia):

(i) Discriminative classifiers−models of the conditional probability of the target Y, given an observation x, symbolically,$P(Y|X = x)$;

(ii) Generative classifiers−models of the joint probability distribution $P(X, Y)$;

It's intuitive that we can distinguish these two kinds of models through probability theory approaches. In other words, a discriminative model draws the boundaries of classes, while a generative model generates the joint distribution of features and targets.

2 A Story about Hospitals

Given two hospitals A and B, patients in a town can choose to get treatment from either of them. Past information shows that people who go to hospital A has an average recovery rate 60%, while 40% of those who choose B get fully cured.

This morning, you saw your neighbor Robert come back from hospital A. In some way, you can predict that the probability he gets cured is 60%. If we consider another situation, where you need to have an operation, is it a better choice for you to go to hospital A? In other words, do you think the probability that you recover when choosing A will be 20% larger than that when choosing B?

The answer may be 'NO'. An intuitive explanation is that, Robert is your observation while yourself is faced with a random experiment. According to previous definition, Robert is classified in the group where people choose hospital B, so you can use discriminative method to predict his recovery rate. For yourself, however, you are actually in neither A nor B, i.e. you have not made your decision, thus the discriminative method doesn't work. Under this constraint, you may have to find out the causal effects of hospital on the recovery rate. This is why we need the generative model.

3 Derivation Using Probability Theory

In this section, we will introduce the math to illustrate the differences between discriminative and generative models. Recall the definition we have shown before, discriminative classifiers compute $P(Y|X = x)$ while the generative classifiers compute $P(X, Y)$. Now we discuss these two models separately. Assume we observe two random variables X and

Y, If X and Y are discrete, we have:

$$E[Y|X=x] = \sum_y yp(y|X=x) \tag{1}$$

$$E[Y] = E[E[Y|X=x]] = \sum_x \sum_y yp(x,y) \tag{2}$$

$$E[Y] - E[Y|X=x] = \sum_x \sum_y yp(x,y) - \sum_y yp(y|X=x) \tag{3}$$

If X and Y are continuous, we have:

$$E[Y|X=x] = \int_{-\infty}^{\infty} yf_Y(y|x)dy \tag{4}$$

$$E[Y] = E[E[Y|X=x]] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx \tag{5}$$

$$E[Y] - E[Y|X=x] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx - \int_{-\infty}^{\infty} yf_Y(y|x)dy \tag{6}$$

According to the Bayes rule

$$p(y|X=x) = \frac{p(x,y)}{p(x)} \tag{7}$$

$$f(y|x) = \frac{f(x,y)}{f_X(x)} \tag{8}$$

(i) If the distribution of Y is independent of X, i.e.

$$p(x,y) = p(x)p(y) \tag{9}$$

$$f(x,y) = f_X(x)f_Y(y) \tag{10}$$

Then we can derive that:

$$
\begin{aligned}
E[Y] - E[Y|X=x] &= \sum_x \sum_y yp(x,y) - \sum_y yp(y|X=x) \\
&= \sum_x \sum_y yp(x)p(y) - \sum_y y\frac{p(x)p(y)}{p(x)} \\
&= \sum_x p(x) \sum_y yp(y) - \sum_y yp(y) \\
&= \sum_y yp(y) - \sum_y yp(y) = 0
\end{aligned} \tag{11}
$$

Or

$$
\begin{aligned}
E[Y] - E[Y|X=x] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dydx - \int_{-\infty}^{\infty} yf_Y(y|x)dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf_X(x)f_Y(y)dydx - \int_{-\infty}^{\infty} y\frac{f_X(x)f_Y(y)}{f_X(x)}dy \\
&= \int_{-\infty}^{\infty} f_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy - \int_{-\infty}^{\infty} yf_Y(y)dy \\
&= \int_{-\infty}^{\infty} yf_Y(y)dy - \int_{-\infty}^{\infty} yf_Y(y)dy = 0
\end{aligned} \tag{12}
$$

(ii) If the distribution of Y is correlated with X, i.e.

$$p(x,y) \neq p(x)p(y) \tag{13}$$

$$f(x,y) \neq f_X(x)f_Y(y) \tag{14}$$

Then

$$E[Y] \neq E[Y|X=x] \tag{15}$$

Hence, the discriminative method will have a bias on generating the distribution of Y.

Derivation above indicates that, when we want to learn the correlation between two random variables, for instance, the features and targets, the discriminative approach may not work, because we really don't want them to be independent.

4 Advantages and Disadvantages

Based on the derivation of previous section, we can easily get some indications about the pros and cons when applying discriminative or generative approach.

Discriminative Classifiers:

- Usually used to predict the targets when features are given;
- Don't need the information about joint distribution;
- Biased on generating the causal effects.

Generative Classifiers:

- Usually used to generate the causal effects of features on targets;
- Also available on prediction;
- Much more difficult than discriminative approaches.

5 Conclusion

This article provides a preliminary view on discriminative and generative models. We apply case study and probability theory to illustrating the differences between these two approaches, briefly listing their advantages as well as disadvantages. Generally speaking, discriminative models allow you to use conditional information to predict while generative models require adequate data so as to get a complete picture of population.

In practice, there is no absolutely dominant way. The approach you choose depends on what you want to achieve as well as how much information you can collect.

References

[1] Batra, Devesh, Discriminative vs. Generative models, *Github*, May 2, 2016.

[2] Discriminative model & Generative model, *Wikipedia, the free encyclopedia.*

[3] Ross, Sheldon M., *A First Course in Probability, 9th edition, Pearson Education.*

[4] Stock, James H, *Introduction to Econometrics, 3rd edition update*, Pearson Education.