EE698R: Lecture Note 2
Introduction to Generative Models: Sampling and Graphical Models

Upal Rakshit
upalrak20@iitk.ac.in

Arkaprava Biswas
arkapravab20@iitk.ac.in

January 10, 2022

1 Introduction to Probabilistic Models and Use Cases

Until now we have seen classification and regression problem, in which F(x) is as close to as $y_{desired}$ possible where F(x) is output of model F with input x. We model F(x) as the likelihood of the probability $P(y \mid x)$ and correspondingly maximize it to find optimum prediction.

$$F(x) \approx y_{desired}$$
$$F(x) = P(y \mid x)$$
$$\max P(y_{desired} \mid x)$$

Now what if we face other problems like:

1.1 A new set of problems:

1.1.1 eg 1:

Suppose, we've data corresponding to 5 classes, and there's one class of data, which has a lot fewer samples than other classes. In that case, we'd want to generate more samples of that particular prediction to achieve good model performance. But how would we do that.

But here comes two problems:

1. How to sample efficiently from a given distribution p(x) ?

2. Given sample x, how to estimate p(x) and sample more x from it.

1.1.2 eg 2:

Suppose, we've images of faces. Now what if we want to generate new images of those faces with different emotions for some application. How would we do that?

1.1.3 eg 3:

How would we generate new artistic images or new music or new texts by combining different concepts.

1.1.4 eg 4:

How would we differentiate fake samples from real ones? We can do that estimating distribution of real data.

1.1.5 eg 5:

How would we fill in missing values like a missing nose or eye in a face, or a missing certain pixel values of an image, or predict coordinate of a certain pixel value in an image.

1.1.6 eg 6:

Given p(x), find $E_{x\sim p(x)}$[x], $E_{x\sim p(x)}$[x^2], $E_{x\sim p(x)}$[f(x)]

$$E_{x\sim p(x)}[f(x)] = \int_{-\infty}^{+\infty} p(x)\ f(x)\ dx$$

p(x) is complicated, so this integration is not solvable analytically.

Q: How to solve such problem?

A: (1) Generate random samples $\sim$ p(x). (2) Compute sample average/statistics over these samples.

$$X_1, X_2, \ldots, X_N \sim p(x)$$
$$E_{x\sim p(x)}[x] = \int_{-\infty}^{+\infty} x\ p(x)\ dx = \frac{1}{N}\sum_{S=1}^{N} X_S$$
$$E_{x\sim p(x)}[f(x)] = \int_{-\infty}^{+\infty} p(x)\ f(x)\ dx = \frac{1}{N}\sum_{S=1}^{N} f(X_S)$$

1.2 Problems:

Q: What to do if it is difficult to sample from p(x) ?

A: Approximate p(x) with some q(x) which is (1) close to p(x), (2) easy to sample.

Q: How to approximate p(x) ?

A: K-L divergence, it is a metric to compare distributions.

Q: What to do if p(x) is not given but samples are given ?

A: (1) We have to estimate p(x), (2) Then generate more samples from estimated p(x).

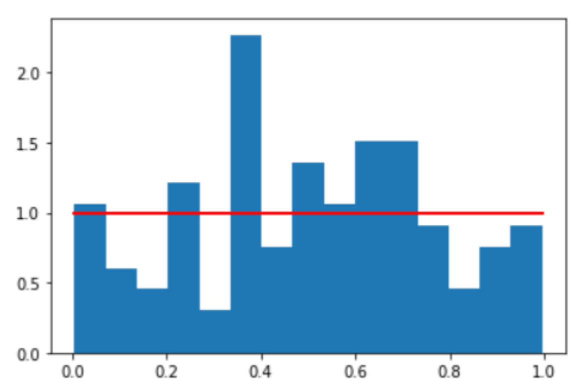
2 Sampling

2.1 What is Sampling ?

Given p(x), where x is a random variable (continuous or discrete). Draw samples of x from p(x).

Uniform random generator (continuous), Z_S = np.random(1)

$Z_S \sim$ U(0,1), $z_1 = 0.412, z_2 = 0.798, z_3 = 0.112\ldots$ 100 samples. It roughly follows uniform distribution because of less number of samples. Histogram plot of the probability density function for 100 samples are shown below.



Q: How to sample non-uniform distributions ?

A: Let a random variable x(discrete); x $\sim$ p(x); p(x) is non-uniform.

S = {H,T}

$$p(x) = \begin{cases} 0.7 & \text{if } x = H \\ 0.3 & \text{if } x = T \end{cases}$$

If x is below 0.7, it is Head otherwise Tail.

Now, let a random variable x(continuous); $z \sim p(z)$; p(z) is non-uniform.
In this case we use transformation of variables. x = f(z)

$$p_x(x) = \sum_{z=f^{-1}(x)} \frac{p_z(z)}{|\frac{df(z)}{dz}|}$$

If f^{-1} is differentiable, then

$$p_x(x) = p_z(z)\, |\frac{df^{-1}(x)}{dx}|; \text{ where } z = f^{-1}(x)$$

If $x = f(z) = z^2$

$$p_x(x) = \frac{p_z(z_1)}{|2z_1|} + \frac{p_z(z_2)}{|2z_2|}$$

$$p_x(x) = \frac{p_z(\sqrt{x})}{2\sqrt{x}} + \frac{p_z(-\sqrt{x})}{2\sqrt{x}}$$

For any $p_x(x)$ and $p_z(z) \sim U(0,1)$, we use

$$x = F_x^{-1}(z)$$

$$\text{CDF, } F_x(x) = \int_{-\infty}^{x} p_x(x)\, dx$$

$$p_x(x) = p_z(z)\, |\frac{df^{-1}(x)}{dx}|; \text{ where } z = f^{-1}(x)$$

Here, $f = F^{-1}$

$$p_x(x) = p_z(z)\, |\frac{d(F^{-1})^{-1}(x)}{dx}|$$

Now, $p_z(z) \sim U(0,1)$

$$|\frac{dF(x)}{dx}| = p_x(x) \ldots\text{[Proved]}$$

2.2 Multivariate Sampling:

Q: How to do sampling if x is a vector ?

A: Here comes the concept of Multivariate sampling. There are two methods of multivariate sampling, (1) By Factorizing and (2) By Transforming

2.2.1 By Factorizing:

$$x = (x_1, x_2) \sim p_{x_1,x_2}(x_1, x_2)$$

x_1 - Ball shape index; x_2 - Basket index

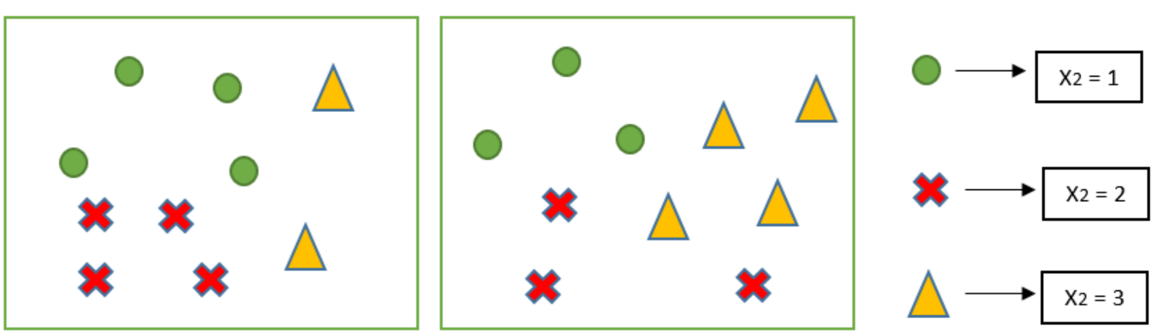


Fig: X1 = 1

Fig: X1 = 2

$$p_{x_1}(x_1) = \begin{cases} 0.7 & \text{if } x_1 = 1 \\ 0.3 & \text{if } x_1 = 2 \end{cases}$$

$$p_{x_1,x_2}(x_1, x_2) = p_{x_1}(x_1)\ p_{x_2|x_1}(x_2|x_1)$$

$$p_{x_2|x_1}(x_2|x_1)$$

	x_2=1	x_2=2	x_2=3
x_1=1	5/11	4/11	2/11
x_1=2	4/11	3/11	4/11

For sampling, sample $x_1 \sim p_{x_1}(x_1)$, $x_2 \sim p_{x_2|x_1}(x_2|x_1)$
Number of parameters used = (2-1) + 2*(3-1) = 5

2.2.2 By Transformation:

$$Z \sim p_{z_1}(z_1)\ p_{z_2}(z_2)$$
$$x = f(z)$$
$$Z \sim N(Z;0,I)$$

Z is easy to sample.

Let $x \sim N(x;\mu,\Sigma)$
take $x = \Sigma Z + \mu$
Now, we can sample Z and obtain x.

3 Graphical Models

Let scalar random variables $x_1, x_2, \ldots, x_7$.
$p(x_1, x_2, \ldots, x_7) = \prod_{n=1}^{7} p(x_n|x_1, x_2, \ldots, x_{n-1}) = p(x_1)\ p(x_2|x_1)\ p(x_3|x_1, x_2)\ldots\ldots$

Q: How many tunable parameters required for defining this distribution ? (Assume for each x_i ϵ 0,1,2,3)
A: Number of parameters = (4-1) + 4*(4-1) + 4^2*(4-1) + …… + 4^6*(4-1) = 4^7-1. Exponentially increasing with respect to random variables.

3.1 Conditional Independence Property:

Conditional independence property can help to reduce the complexity.

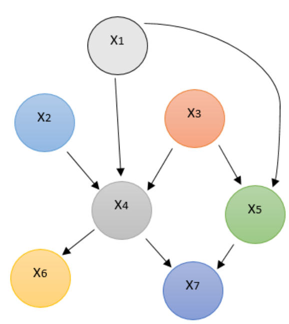


Fig: Directed Acyclic Graph (DAG)

$p(x_1, x_2, \ldots, x_7)$
$= p(x_1)\ p(x_2)\ p(x_3)\ p(x_4|x_1, x_2, x_3)\ p(x_5|x_1, x_3)\ p(x_6|x_4)\ p(x_7|x_4, x_5)$
$= \prod_{n=1}^{7} p(x_n|Parents(x_n))$

Here is no closed loops. It is known as Directed Acyclic Graph(DAG).

4 Reading Assignment

Pattern Recognition and Machine Learning By Christopher M. Bishop:

Sec. 11.0, 11.1.1 - Sampling
Sec. 8.1 - Graphical Modelling
Sec. 11.1.1 - Box-Muller Technique