Generative models
vs
Discriminative models

Discriminative models

A Discriminative model learns the conditional probability distribution $p(y|x)$

A Discriminative model models the decision boundary between the classes.

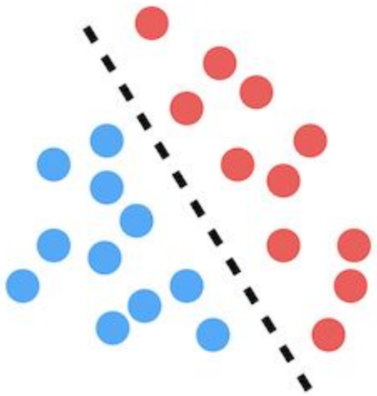


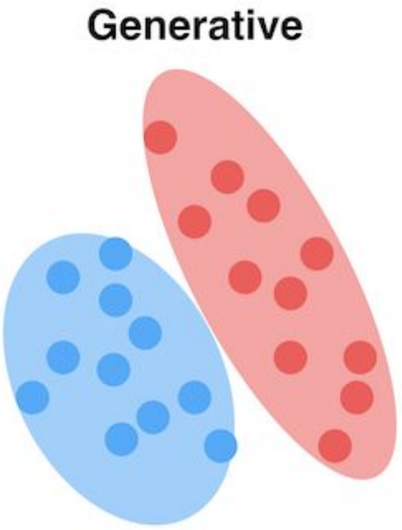
Image source:google image

Generative models

A Generative Model learns the joint probability distribution P(x,y).

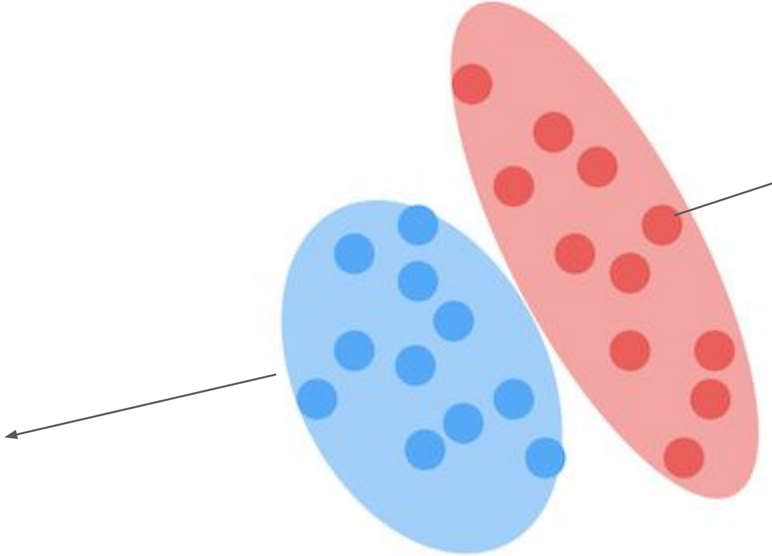
$P(x, y) = P(x \mid y) * P(y)=P(y|x)*P(x)$

A Generative Model explicitly models the actual distribution of each class.



Example: Sheep or Goat?

Goat
model:Y2



Sheep
model:
Y1

P(Y1|X)
P(Y2|X)

If
P(Y1|X)>P(
Y2|X) :
Sheep

Image source:google image

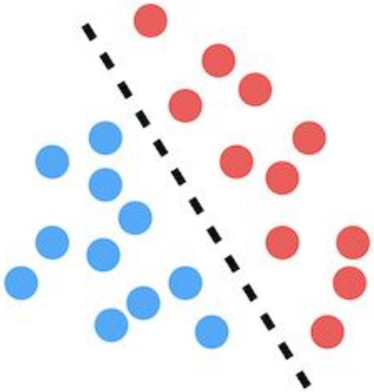


Image source:google image

y=f(x)
->left of
boundary
->sheep

Examples [edit]

Simple example [edit]

Suppose the input data is $x \in \{1, 2\}$, the set of labels for x is $y \in \{0, 1\}$, and there are the following 4 data points: $(x, y) = \{(1,0), (1,1), (2,0), (2,0)\}$

For the above data, estimating the joint probability distribution $p(x, y)$ from the empirical measure will be the following:

	$y=0$	$y=1$
$x=1$	$1/4$	$1/4$
$x=2$	$2/4$	0

p(y|x)*p(x)

while $p(y|x)$ will be following:

	$y=0$	$y=1$
$x=1$	$1/2$	$1/2$
$x=2$	1	0

source：https://en.wikipedia.org/wiki/Generative_model

Bayes rule is better?

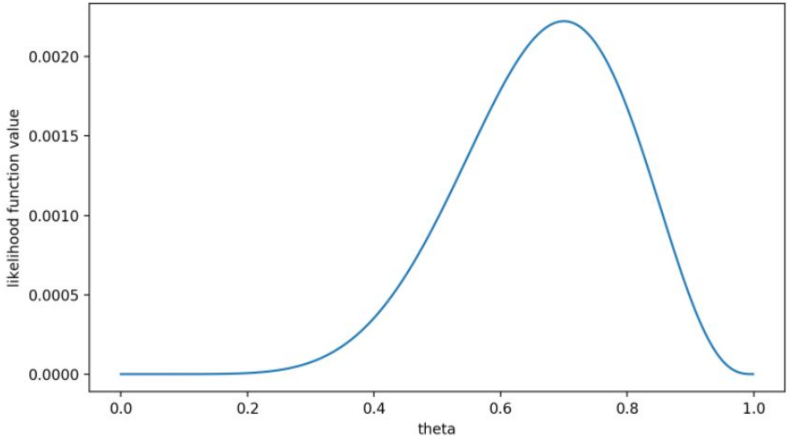
Suppose we toss coin 10 times: THHHHTHHHT

Maximum likelihood estimation

Θ is the probability that a H occur.

X0 is our 10-times-toss experiment

$f(x0,\theta)=(1-\theta)\times\theta\times\theta\times\theta\times\theta\times(1-\theta)\times\theta\times\theta\times\theta\times(1-\theta)=\theta^7*(1-\theta)^3=f(\theta)$ -> Likelihood Function



Based on the result we think the probability that we get a head is 0.7.

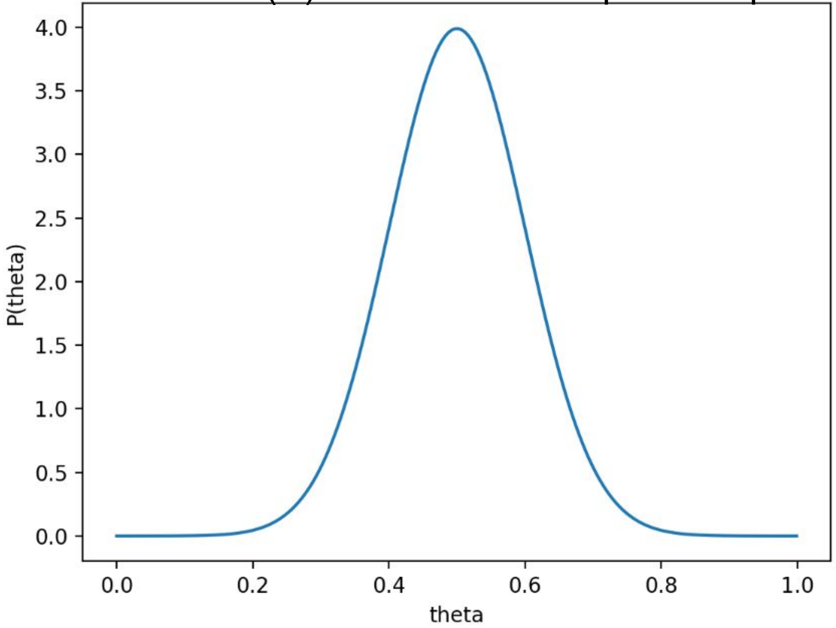
However, no one would be convinced. All of us know that it should be 0.5…..

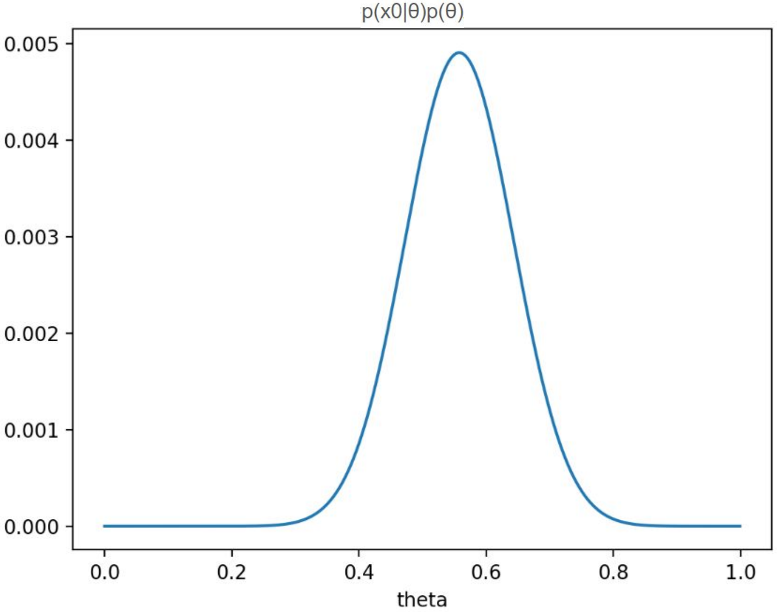
source:https://blog.csdn.net/u011508640/article/details/7

Maximum a posteriori estimation

P(θ | x0)=p(x0|θ)p(θ)/p(x)-> Bayes Rule

A Priori P(θ) → 0.5 as our empirical experiment

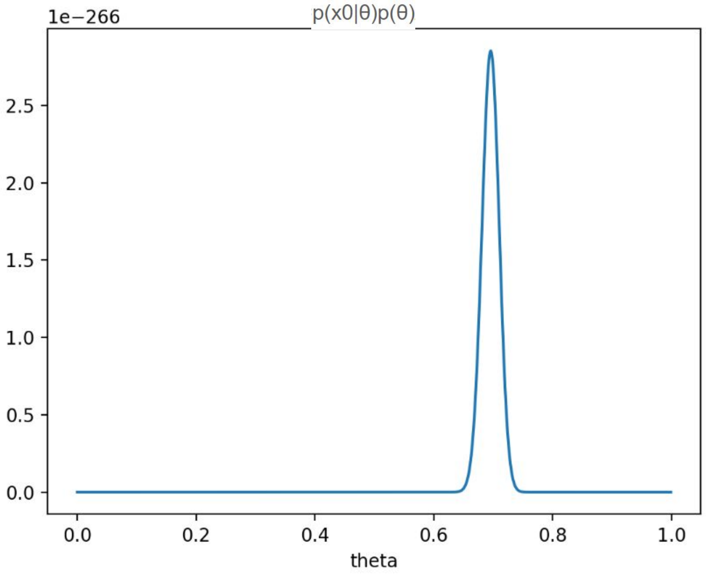




θ=0.558 in this case which is close to a priori or empirical experience.

What if our coin has some problem that the probability is really 0.7 for a head?

We need to do more experiments. If we do our experiment 1000 times and we get 700 heads.



When θ=0.696 we get the maximum value. Finally, even we have empirical experiment, we still believe that the probability is 0.7 for a head