PARIS-EAST CRÉTEIL UNIVERSITY

INTERNSHIP THESIS

Classifications of Poisson processes and spatial point Poisson processes

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MATHEMATICS AND COMPUTER SCIENCE MASTER 2 TRACK

Author:
Mamadou DIOUF
Jury:
Pr Laurent HAUSWIRTH
Pr Ciryl NICAUD

Supervisor:
MCF Miguel MARTINEZ
Advisor:
MCF Mohamed HEBIRI
Referent teacher:
Pr Vincent MILLOT

September 4, 2025

Description of the internship structure

Laboratoire d'Analyse et de Mathématiques Appliquées (LAMA) is a joint research unit (UMR 8050) affiliated with three institutions: CNRS, the University Paris-Est Créteil (UPEC), and the Gustave Eiffel University (UGE). Its premises are located on two sites: one in Champs-sur-Marne, within UGE, and the other in Créteil, within UPEC. LAMA has over a hundred members, including about 70 permanent researchers and faculty members. LAMA is a member of the Labex Bézout "Models and Algorithms: from Discrete to Continuous" and the CNRS Bézout Research Federation (FR 3522). It is also a member of the Graduate Program SFRI Bézout.

Its Probability and Statistics team consists of just under thirty permanent members. Their research themes include (but are not limited to):

- Limit theorems: Several members of the team study the asymptotic behavior of sums of random variables, Markov chains, or the spectrum of large random matrices.
- Probabilistic numerical methods: Research activities focus on numerical schemes for Markov processes, including SDEs (Stochastic Differential Equations), SDERs (Stochastic Differential Equations with Rough paths), and ESDEs (Exponential Stochastic Differential Equations).
- Theory of stochastic processes: Various questions in the theory of processes are addressed, such as density estimation for processes, existence problems for SDEs/SDERs, stochastic control, or piecewise deterministic Markov processes.
- Mathematical finance: The finance component of the team focuses on option hedging strategies, optimal investment, counterparty costs, exponential Lévy models, American options, and high-frequency trading.
- Stochastic models: The team works on stochastic models arising from questions in the life sciences, considering both their practical applications and their purely theoretical aspects, especially models related to multitype branching processes.
- Statistical mechanics: Several members of the team work on polymers, diffusion-limited aggregation, Z-invariance, the Ising model, metastability, Gibbs measures, two-dimensional quantum gravity, and the quantum theory of Liouville.
- Statistics: The team is interested in copulas for order statistics, confidence sets in supervised learning, population mixture models, parameter estimation for evolutionary models, non-parametric tests for large covariance matrices, classification, extreme value distributions, high-dimensional statistics, quantum statistics, and estimation of jumps in stochastic processes.

Dedication

I would like to dedicate this thesis to the remarkable individuals who have supported and inspired me throughout my academic journey.

To my mathematics teachers, whose passion for the subject and commitment to excellence have not only expanded my knowledge but also instilled in me a deep love for learning. Your guidance has been instrumental in shaping the academic path that has led me here.

To my family, whose unwavering support, love, and encouragement have been my constant source of strength. Your belief in me has fueled my determination, and your sacrifices have made every achievement possible. This thesis is as much yours as it is mine.

Contents

Acknowledges

I would like to express my sincere gratitude to my supervisor, MCF Miguel MARTINEZ for his support, valuable feedback, and guidance throughout the course of my internship and research. His expertise has been a vital part of my learning experience. I would like to extend my sincere gratitude to MCF Mohamed HEBIRI, for your dedicated assistance during our two extend work sessions.

A special thanks to the Labex Bézout, for providing me with the opportunity to grow, learn, and contribute. Their program have been pivotal in my development.

I am deeply grateful to Paris-Est Créteil and Gustave Eiffel universities for providing the academic platform that allowed me to undertake this internship and research. The knowledge and resources provided by the institution have been fundamental in completing this work.

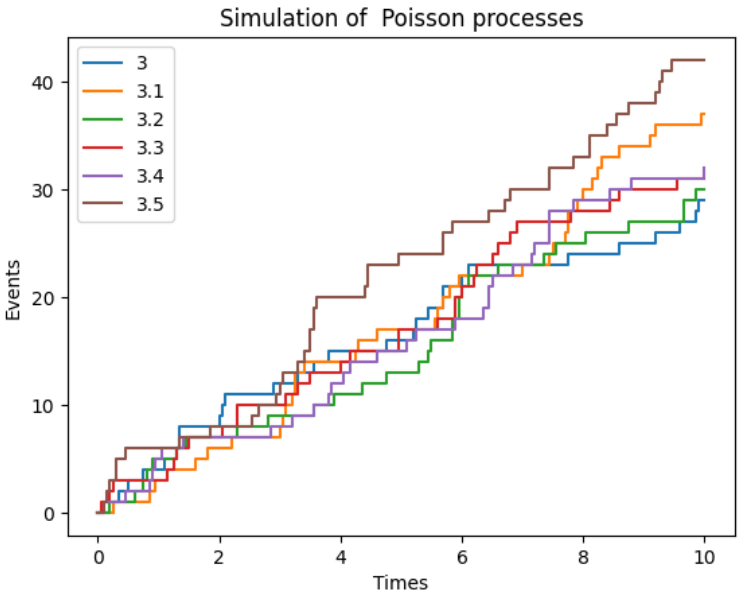
My sincere thanks also go to my family and friends for their constant love, support, and encouragement. Their belief in me has been a source of motivation, and I am truly blessed to have them by my side.

1 Abstract

In this theses, we will provide the definitions and the properties of a homogeneous Poisson process, his inter-arrival time distribution and his arrival time distribution. And the classification problem in machine learning specially the Bayes classifier will be provided with an application on Poisson processes. A classification based on the parameter λ of the Poisson process.

2 Introduction

A Poisson process is a mathematical model used to describe random events that occur independently over time or space at a constant average rate. It is characterized by a single parameter, λ (lambda), which represents the average rate of events per unit of time (or space). In machine learning, classification is a supervised learning task where the goal is to predict the categorical class label for new, unseen data points based on patterns learned from a labeled dataset. Essentially, we're teaching a computer to categorize things. Our goal is to categorize the Poisson processes by there parameters λ, the labels, find the better λ for new Poisson process in our labels.



3 Poisson processes

3.1 Point processes

Definition 1: Stochastic process

A stochastic process is a series (collection) of random variables that describes the evolution of a random phenomenon generally with respect to time. A stochastic process can be discrete-time denoted as

$$\{X_n; n \geq 0\}$$

or continuous-time denoted as

$$\{X(t); t \geq 0\}$$

Definition 2: Point process

A point process is a stochastic process with a series of random variables that represent the arrival times (waiting times) of an event, and can be denoted as

$$\{T_n; n \geq 0\}$$

where T_n: Arrival time (waiting time) of the n^{th} event, n^{th} arrival time

$$0 \leq T_0 < T_1 < \cdots < T_n$$

Definition 3: Inter-arrival time

Consider the point process $\{T_n; n \geq 0\}$ where T_n is the arrival time of the n^{th} event. Then

$$I_n = T_n - T_{n-1}\ ; n \geq 1 \text{ is the } n^{th} \text{ inter-arrival time}$$

$\{I_n; n \geq 1\}$ is a stochastic process with the random variables denoting the inter-arrival times. The stochastic process of the inter-arrival times . And we have

$$T_n = I_1 + I_2 + \cdots + I_n$$

3.2 Poisson process

Definition 4: Counting process

A counting process is a stochastic process with a series of random variables that represent the number of events in a given time interval, and can be denoted as

$$\{N(t); t \geq 0\}$$

where $N(t)$: Number of events in the time interval $(0, t]$.

It is assumed that $N(0) = 0$, i.e. at $t = 0$, no event occurs. Note that

$$N(t) - N(s) : \text{Number of events in the time interval } (s, t] \text{ for } s < t$$

Definition 5: Stationary increments and independent increments of a counting process

Stationary increments occur if the distribution of the number of events depends only on the length of the time interval. As an example, in case of stationary increments, $N(t+s) - N(s)$ will have the same distribution with $N(t)$ for all $s \geq 0,\ t > 0$, since the length of the time interval is t for both increments.

Independent increments occur if the number of events in disjoint time intervals is independently distributed. As an example, in case of independent increments,
$N(t) - N(s)$ and $N(s)$ will be independently distributed for all $s \geq 0,\ t > 0$, since $(0, s]$ and $(s, t]$ are disjoint intervals .

Definition 6.1: Homogeneous Poisson Process

An homogeneous Poisson Process with rate $\lambda,\ \lambda > 0$ is counting process $\{N(t); t \geq 0\}$ which satisfies :

- $N(0) = 0$
- $\{N(t); t \geq 0\}$ are independent increments : $\forall n\ 0 \leq t_0 < t_1 < \cdots < t_n$,

  $N(t_1) - N(t_0), N(t_2) - N(t_1), \cdots, N(t_n) - N(t_{n-1})$ are independently distributed.
- Stationary increments $\forall t \geq 0, \forall s > 0$

  $$P(N(t+s) - P(N(s) = n) = P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

  $N(t+s) - N(t)$ will have the same distribution with $N(t)$ for all $s \geq 0, t > 0$, since the length of the time interval is t for both increments.

Definition 6.2: Homogeneous Poisson Process

A counting process $\{N(t); t \geq 0\}$ is said to be a Poisson process with rate $\lambda,\ \lambda > 0$, if

$$N(t) = \sum_{n \geq 1} \mathbb{1}_{(T_n \leq t)} \tag{1}$$

for a sequence $(T_n)_{n \geq 1}$ with $Gamma(n, \lambda)$ having i.i.d. increments $(I_n)_{n \geq 1}$ with $Exponential(\lambda)$ distribution.

Proposition 1: Let $A \in \mathbb{R}^n$, $T_0 = 0$ and $T_n = \sum_{i=1}^{n} I_i$, $n \in \mathbb{N}^*$ then $\forall\, t \geq 0$, $N(t) = \sum_{n\geq 1} \mathbb{1}_{(T_n \leq t)}$, we have:

1. $(I_1, ..., I_n) \sim \lambda^n e^{-\lambda \sum_{k=1}^{n} i_k}$
2. $$\mathbb{P}((I_1, ..., I_n) \in A, N(t) = n) = \int_A e^{-\lambda t} \lambda^n \mathbb{1}_{i_1 + i_2 + \cdots + i_n \leq t} di_1 di_2 \cdots di_n \tag{2}$$
3. $(T_1, ..., T_n) \sim \lambda^n e^{-\lambda t_n} \mathbb{1}_{0 < t_1 < \cdots < t_n}$.
4. $$\mathbb{P}((T_1, ..., T_n) \in A, N(t) = n) = \int_A e^{-\lambda t} \lambda^n \mathbb{1}_{0 < t_1 < t_2 < \cdots < t_n \leq t} dt_1 dt_2 \cdots dt_n \tag{3}$$

Proof:

1. $(I_k), k \in [\![1; n]\!]$ are (independent and identically distributed) i.i.d .
2. We have $N(t) = n \iff (T_n \leq t < T_{n+1})$, $N(t)$ a Poisson process with rate λ and $t_j = i_1 + \cdots + i_j$.

$$\begin{aligned}
\mathbb{P}((I_1, \ldots, I_n) \in A, N(t) = n) &= \mathbb{P}((I_1, \ldots, I_n) \in A, T_n \leq t < T_{n+1}) \\
&= \iint_{A \times \mathbb{R}_+^*} \mathbf{1}_{i_1 + i_2 + \cdots + i_n \leq t < t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} di_1 \ldots di_n di_{n+1} \\
&= \int_A \mathbf{1}_{t_n \leq t} \lambda^n e^{-\lambda t_n} di_1 \ldots di_n \left(\int_{t - t_n}^{+\infty} \lambda e^{-\lambda i_{n+1}} di_{n+1} \right) \\
&= \int_A e^{-\lambda t} \lambda^n \mathbf{1}_{t_n \leq t} di_1 \ldots di_n.
\end{aligned}$$

3. we have $N(t) = n \iff (T_n \leq t < T_{n+1})$, and $N(t)$ a Poisson process with rate λ. Let h be a Borel measurable function positive definite in $\mathbb{R}^n$. Since the random variable $I_1, \ldots, I_n$ are independents, we have

$$E(h(T_1, \ldots, T_n)) = \int_D h(i_1, i_1 + i_2, \ldots, i_1 + \cdots + i_n) \lambda^n e^{-\lambda i_n} di_1 \ldots di_n,$$

where $D \subset (\mathbb{R}_+^*)^n$. By making the change of variable $t_j = i_1 + \cdots + i_j$, $j = 1, \ldots, n$, we obtain

$$E(h(T_1, \ldots, T_n)) = \int_{\mathbb{R}^n} h(t_1, \ldots, t_n) \lambda^n e^{-\lambda t_n} \mathbf{1}_{0 < t_1 < \cdots < t_n} dt_1 \ldots dt_n.$$

Hence the density of $(T_1, \ldots, T_n)$ is : $\lambda^n e^{-\lambda t_n} \mathbf{1}_{0 < t_1 < \cdots < t_n}$.

4. We have $P(N(t)=n)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}$ and $N(t)=n \iff T_n \le t < T_{n+1}$. By the result of 1 we obtain,

$$\begin{aligned}
\mathbb{P}((T_1,\dots,T_n)\in A, N(t)=n) &= \mathbb{P}((T_1,\dots,T_n)\in A, T_n\le t<T_{n+1}) \\
&= \iint_{\mathbb{A}\times\mathbb{R}} \mathbf{1}_{0<t_1<\cdots<t_n\le t<t_{n+1}} \lambda^{n+1} e^{-\lambda t_{n+1}} dt_1\dots dt_n dt_{n+1} \\
&= \int_A \mathbf{1}_{0<t_1<\cdots<t_n\le t} \lambda^n dt_1\dots dt_n \left(\int_t^{+\infty} \lambda e^{-\lambda t_{n+1}} dt_{n+1}\right) \\
&= \int_A e^{-\lambda t}\lambda^n \mathbf{1}_{0<t_1<\cdots<t_n\le t} dt_1\dots dt_n.
\end{aligned}$$

Hence

$$\mathbb{P}((T_1,\dots,T_n)\in A, N(t)=n) = \int_A e^{-\lambda t}\lambda^n \mathbf{1}_{0<t_1<\cdots<t_n\le t} dt_1\dots dt_n.$$

[3] □

Algorithm 1: Simulation of Poisson process

Simulation of Poisson process

Algorithm 1 Poisson Process Simulation Algorithm

Require: λ (lambda): The average rate of events (events per unit of time).
Require: $total\_time$: The total duration of the simulation.
Ensure: $arrival\_times$: A list of the times at which events occur.
Ensure: $time\_event\_counts$: A tuple containing two lists:
$times$: A list of time points.
$event\_counts$: A list of the number of events that have occurred up to each time point in $times$.

$arrival\_times \leftarrow []$
$current\_time \leftarrow 0$
while $current\_time < total\_time$ **do**
    Generate $inter\_arrival\_time$ from an exponential distribution with mean $1/\lambda$.
    $current\_time \leftarrow current\_time + inter\_arrival\_time$
    if $current\_time < total\_time$ **then**
        Append $current\_time$ to $arrival\_times$
    end if
end while
Generate a list of time points, $times$, from 0 to $total\_time$.
Create an empty list $event\_counts$.
for each $time$ in $times$ **do**
    $count \leftarrow$ number of $arrival\_times$ that are $\leq time$
    Append $count$ to $event\_counts$
end for
return $arrival\_times$ and $(times, event\_counts)$

3.3 Poisson point processes

Poisson point process (also known as: Poisson random measure, Poisson random point field and Poisson point field) with rate λ is a type of mathematical object that consists of points randomly located on a mathematical space with the essential feature that the points occur independently of one another. If the rate λ is constant, it is the average density of the points in the Poisson process located in some region of space. The resulting point process is called a homogeneous or stationary Poisson point process. Otherwise the point process is called an inhomogeneous or nonhomogeneous Poisson

point process, and the average density of points depend on the location of the underlying space of the Poisson point process.

3.3.1 Homogeneous Poisson point process

Real line Poisson point process

A Poisson point process can be defined on the real line by considering the number of points of the process in the interval $(a,b]$. For the homogeneous Poisson point process on the real line with parameter $\lambda > 0$, the probability of this random number of points, written here as $N(a,b]$, being equal to some counting number n is given by:

$$\mathbb{P}\big(N(a,b] = n\big) = \frac{[\lambda(b-a)]^n}{n!} e^{-\lambda(b-a)},$$

For some positive integer k, the homogeneous Poisson point process has the finite-dimensional distribution given by:

$$\mathbb{P}\big(N(a_i,b_i] = n_i, i = 1,\ldots,k\big) = \prod_{i=1}^{k} \frac{[\lambda(b_i-a_i)]^{n_i}}{n_i!} e^{-\lambda(b_i-a_i)},$$

where the real numbers $a_i < b_i \leq a_{i+1}$.

Spatial Poisson point process

A spatial Poisson process is a Poisson point process defined in $\mathbb{R}^d$. For some bounded region S of Euclidean space $\mathbb{R}^d$, if the points form a homogeneous Poisson process with parameter $\lambda > 0$, then the probability of n points existing in $S \subset \mathbb{R}^d$ is given by:

$$\mathbb{P}\big(N(S) = n\big) = \frac{(\lambda|S|)^n}{n!} e^{-\lambda|S|}$$

where $|S|$ now denotes the d-dimensional volume of S. Furthermore, for a collection of disjoint, bounded Borel sets $S_1,\ldots,S_k \subset \mathbb{R}^d$, let $N(S_i)$ denote the number of points of N existing in S_i. Then the corresponding homogeneous Poisson point process with parameter $\lambda > 0$ has the finite-dimensional distribution:

$$\mathbb{P}\big(N(S_i) = n_i, i = 1,\ldots,k\big) = \prod_{i=1}^{k} \frac{(\lambda|S_i|)^{n_i}}{n_i!} e^{-\lambda|S_i|}.$$

Positioning of points

Homogeneous Poisson point processes do not depend on the position of the underlying space through its parameter λ, which implies it is both a stationary process (invariant to translation) and an isotropic (invariant to rotation) stochastic process.

If the homogeneous point process is defined on the real line as a mathematical model for occurrences of some phenomenon, then it has the characteristic that the positions of these occurrences or events on the real line (often interpreted as time) will be uniformly distributed. This uniformity property extends to higher dimensions in the Cartesian coordinate, but not in, for example, polar coordinates.

3.3.2 Inhomogeneous Poisson point process

The inhomogeneous or nonhomogeneous Poisson point process is a Poisson point process with a Poisson parameter set as some location-dependent function in the underlying space on which the Poisson process is defined. For Euclidean space $\mathbb{R}^d$, this is achieved by introducing a locally integrable positive function $\lambda\colon \mathbb{R}^d \to [0,\infty)$, such that for every bounded region S the (d-dimensional) volume integral of $\lambda(x)$ over region S is finite. In other words, if this integral, denoted by $\Lambda(S)$, is:

$$\Lambda(S) = \int_S \lambda(x)\,\mathrm{d}x < \infty,$$

where $\mathrm{d}x$ is a (d-dimensional) volume element,then for every collection of disjoint bounded Borel measurable sets $S_1,\ldots,S_k$, an inhomogeneous Poisson process with (intensity) function $\lambda(x)$ has the finite-dimensional distribution:

$$\mathbb{P}\big(N(S_i) = n_i, i = 1,\ldots,k\big) = \prod_{i=1}^{k} \frac{(\Lambda(S_i))^{n_i}}{n_i!} e^{-\Lambda(S_i)}.$$

Furthermore, $\Lambda(S)$ has the interpretation of being the expected number of points of the Poisson process located in the bounded region S, namely $\Lambda(S) = \mathrm{E}[N(S)]$.

Real line Poisson point process

On the real line, the inhomogeneous or non-homogeneous Poisson point process has mean measure given by a one-dimensional integral. For two real numbers a and b, where $a \le b$, denote by $N(a,b]$ the number points of an inhomogeneous Poisson process with intensity function $\lambda(t)$ occurring in the interval $(a,b]$. The probability of n points existing in the above interval $(a,b]$ is given by:

$$\mathbb{P}\big(N(a,b] = n\big) = \frac{[\Lambda(a,b)]^n}{n!} e^{-\Lambda(a,b)}.$$

where the mean or intensity measure is:

$$\Lambda(a,b) = \int_a^b \lambda(t)\,\mathrm{d}t,$$

which means that the random variable $N(a,b]$ is a Poisson random variable with mean $\mathrm{E}[N(a,b]] = \Lambda(a,b)$

Spatial Poisson point process

An inhomogeneous Poisson process defined in the plane $\mathbb{R}^2$ is called a spatial Poisson process It is defined with intensity function and its intensity measure is obtained performing a surface integral of its intensity function over some region. For example, its intensity function (as a function of Cartesian coordinates x and y) can be

$$\lambda(x,y) = e^{-(x^2+y^2)},$$

so the corresponding intensity measure is given by the surface integral

$$\Lambda(S) = \int_S e^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y,$$

where S is some bounded region in the plane $\mathbb{R}^2$.

In higher dimensions, in the plane,$\Lambda(S)$ corresponds to a surface integral while in $\mathbb{R}^d$ the integral becomes a (d-dimensional) volume integral.

Positioning of points

A couple of different methods can be used depending on the nature of the intensity function $\lambda(x)$. If the intensity function is sufficiently simple, then independent and random non-uniform (Cartesian or other) coordinates of the points can be generated. For example, simulating a Poisson point process on a circular window can be done for an isotropic intensity function (in polar coordinates r and θ), implying it is rotationally variant or independent of θ but dependent on r, by a change of variable in r if the intensity function is sufficiently simple.

For more complicated intensity functions, one can use an acceptance-rejection method, which consists of using (or 'accepting') only certain random points and not using (or 'rejecting') the other points, based on the ratio:

$$\frac{\lambda(x_i)}{\Lambda(S)} = \frac{\lambda(x_i)}{\int_S \lambda(x)\,\mathrm{d}x.}$$

where x_i is the point under consideration for acceptance or rejection.

That is, a location is uniformly randomly selected for consideration, then to determine whether to place a sample at that location a uniformly randomly drawn number in $[0,1]$ is compared to the probability density function $\frac{\lambda(x)}{\Lambda(S)}$, accepting if it is smaller than the probability density function, and repeating until the previously chosen number of samples have been drawn.

Inhomogeneous Poisson process in a separable and bounded metric space

Let (S,ρ) be a separable and bounded metric space, endowed with a Borel measure ν, let us denote by $\mathscr{B}(S)$ the Borel σ-algebra on S and by S^∞ the set of elements (subsets) x of S whose cardinality, $\#x$, is finite. That is,

$$S^\infty = \{x \subset S : \#x < \infty\}.$$

Let $\lambda : S \to \mathbb{R}^+$ be an integrable function. Given a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, we will say that a function $X : \Omega \to S^\infty$ is a *Poisson process on S with intensity* λ (we will denote $X \sim \mathscr{P}(S,\lambda)$) if:

- the functions $N_A : \Omega \to \{0,\dots,\infty\}$ defined by $N_A(\omega) = \#\{\omega : X(\omega) \cap A\}$ are random variables for all $A \in \mathscr{B}(S)$;
- given n disjoint Borel subsets $A_1,\dots,A_n$ of S, the random variables $N_{A_1},\dots,N_{A_n}$ are independent;
- N_A follows a Poisson process with mean $\mu(A)$ (we will write $N_A \sim \mathscr{P}(\mu(A))$), with

$$\mu(A) = \int_A \lambda(\xi)\, d\nu(\xi).$$

Let $S^\infty = 2^S$ be the σ-algebra of part of S^∞. If X is a Poisson process, the distribution $\mathbb{P}_X$ of X on S^∞ is defined by $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$ for $B \in S^\infty$.

for all $X \sim \mathscr{P}(S,\lambda)$, its density is

$$f_X(x) = \exp\{\nu(S) - \mu(S)\} \prod_{\xi \in x} \lambda(\xi)$$

where $\mu(S) = \int_S \lambda\, d\nu$.

4 Bayesian statistics

Bayesian statistics is a theory in the field of statistics based on the Bayesian interpretation of probability, where probability expresses a degree of belief in an event. The degree of belief may be based on prior knowledge about the event, such as the results of

previous experiments, or on personal beliefs about the event. It use Bayes' theorem to compute and update probabilities after obtaining new data. Bayes' theorem describes the conditional probability of an event based on data as well as prior information or beliefs about the event or conditions related to the event.

4.1 Bayes' theorem

Bayes's theorem is used in Bayesian methods to update probabilities, which are degrees of belief, after obtaining new data. Given two events A and B, the conditional probability of A given that B is true is expressed as follows:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

where $\mathbb{P}(B) \neq 0$. Although Bayes's theorem is a fundamental result of probability theory, it has a specific interpretation in Bayesian statistics. In the above equation, A usually represents a proposition (such as the statement that a coin lands on heads fifty percent of the time) and B represents the evidence, or new data that is to be taken into account (such as the result of a series of coin flips).

$\mathbb{P}(A)$ is the prior probability of A which expresses one's beliefs about A before evidence is taken into account. The prior probability may also quantify prior knowledge or information about A.

$\mathbb{P}(B \mid A)$ is the likelihood function, which can be interpreted as the probability of the evidence B given that A is true. The likelihood quantifies the extent to which the evidence B supports the proposition A.

$\mathbb{P}(A \mid B)$ is the posterior probability, the probability of the proposition A after taking the evidence B into account. Essentially, Bayes's theorem updates one's prior beliefs $\mathbb{P}(A)$ after considering the new evidence B. The probability of the evidence $\mathbb{P}(B)$ can be calculated using the law of total probability. If $\{A_1, A_2, \ldots, A_n\}$ is a partition of the sample space, which is the set of all outcomes of an experiment, then,

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \mathbb{P}(B \mid A_2)\mathbb{P}(A_2) + \cdots + \mathbb{P}(B \mid A_n)\mathbb{P}(A_n) = \sum_i \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)$$

When there are an infinite number of outcomes, it is necessary to integrate over all outcomes to calculate $\mathbb{P}(B)$ using the law of total probability. Often, $\mathbb{P}(B)$ is difficult to calculate as the calculation would involve sums or integrals that would be time-consuming to evaluate, so often only the product of the prior and likelihood is con-

sidered, since the evidence does not change in the same analysis. The posterior is proportional to this product:

$$\mathbb{P}(A \mid B) \propto \mathbb{P}(B \mid A)\mathbb{P}(A)$$

The maximum a posteriori, which is the mode of the posterior and is often computed in Bayesian statistics using mathematical optimization methods, remains the same. The posterior can be approximated even without computing the exact value of $\mathbb{P}(B)$ with methods such as Markov chain Monte Carlo or variational Bayesian methods.

4.2 Maximum Likelihood Estimation (MLE)

The maximum likelihood estimation (MLE) is a method of estimating the parameters of an assumed probability distribution, given some observed data. This is a maximum of the likelihood function under the assumed statistical model, where the observed data is most probable.

Let $\mathbb{X}$ be our observed data and λ our parameter in the space $\mathscr{Y}$. if $\mathbb{X}|\lambda$ has a:

- Discrete distribution the likelihood function is $L(\lambda|\mathbb{X}) = \mathbb{P}(X|\lambda)$ and its MLE is $\arg\max_{\lambda\in\mathscr{Y}} L(\lambda|\mathbb{X})$.
- Continuous distribution the likelihood function is $L(\lambda|\mathbb{X}) = f(X|\lambda)$ and its MLE is $\arg\max_{\lambda\in\mathscr{Y}} L(\lambda|\mathbb{X})$.

4.3 Maximum A Posteriori (MAP) estimation

Let $\mathbb{X}$ be our observed data, λ our parameter in the space $\mathscr{Y}$ with $\pi(\lambda)$ its probability distribution. if $\pi(\lambda)$ is a:

- Discrete distribution the posterior probability is : $\pi(\lambda|\mathbb{X}) = \dfrac{L(\lambda|\mathbb{X})\pi(\lambda)}{\sum_{\lambda'\in\mathscr{Y}} L(\lambda'|\mathbb{X})\pi(\lambda')}$

  and its MAP estimation is $\arg\max_{\lambda\in\mathscr{Y}} \pi(\lambda|\mathbb{X})$.

- Continuous distribution the posterior probability is : $\pi(\lambda|\mathbb{X}) = \dfrac{L(\lambda|\mathbb{X})\pi(\lambda)}{\int_{\lambda'\in\mathscr{Y}} L(\lambda'|\mathbb{X})\pi(\lambda')d\lambda'}$

  and its MAP estimation is $\arg\max_{\lambda\in\mathscr{Y}} \pi(\lambda|\mathbb{X})$.

5 Classification

Classification is supervised learning problem, its goal is to learn a mapping from inputs x to outputs y, where $y \in \{\lambda_1, ..., \lambda_J\}$, with J being the number of classes. If $J = 2$, this is called binary classification (in which case we often assume $y \in \{0, 1\}$); if $J > 2$, this is called multiclass classification. If the class labels are not mutually exclusive (e.g., somebody may be classified as tall and strong), we call it multi-label classification, but this is best viewed as predicting multiple related binary class labels (a so-called multiple output model).

One way to formalize the problem is as function approximation. We assume $y = g(x)$ for some unknown function g, and the goal of learning is to estimate the function g given a labeled training set, and then to make predictions using $\hat{y} = \hat{g}(x)$. (We use the hat symbol to denote an estimate.) Our main goal is to make predictions on novel inputs, meaning ones that we have not seen before (this is called generalization), since predicting the response on the training set is easy (we can just look up the answer).

5.1 Prediction algorithm and loss function

A prediction function is a (measurable) function $g : \mathscr{X} \to \mathscr{Y}$. Let $\mathscr{F}(\mathscr{X}, \mathscr{Y})$ be the set of all prediction functions, $Z_i = (X_i, Y_i) \sim P$. $\{Z_1, ..., Z_n\}$ is the training set.

An algorithm is any function that takes a training set to produce a prediction:

$$G : \cup_{n=1}^{\infty} \mathscr{Z}^n \to \mathscr{F}(\mathscr{X}, \mathscr{Y}); \qquad \mathscr{Z} = (\mathscr{X}, \mathscr{Y}).$$

Let $l(y, y')$ be the resulting loss when the true label is y and the predicted label y'. The function $l : \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}$ is the cost function.

5.2 Bayes predictor

The quality of a prediction function $g : \mathscr{X} \to \mathscr{Y}$ is measured through its risk (or generalization error):

$$R_P(g) = \mathbf{E}[l(Y, g(X))] = \int_{\mathscr{X},\mathscr{Y}} l(y, g(x)) d\mathbb{P}(x, y)$$

The "best" prediction function is the (or more precisely, a) function of $\mathscr{F}(\mathscr{X}, \mathscr{Y})$ minimizing the risk R_P:

$$g^* \in \arg \min_{g \in \mathscr{F}(\mathscr{X},\mathscr{Y})} R_P(g)$$

Such a function g^* might not exist but exits for usual loss functions. The "best" function is called $Bayes\ predictor$. It depends on the unknown probability P and is the unknown itself.

5.3 Bayes classifier

Proposition 2: The Bayes classifier $\arg\max_{y\in\mathscr{Y}} \mathbb{P}\big(Y=y\big|X\big)$, the MAP estimation of y, is a Bayes predictor when the loss is (0-1 loss) $l(y,y') = \mathbb{1}(y \neq y')$.

Proof:

Compute the Bayes classifier $g^* = \arg\min_{g\in F(\mathscr{X},\mathscr{Y})} R(g)$.

$$R(g) = \mathbb{E}_p\big[\mathbb{1}\big(Y \neq g(X)\big)\big] = \mathbb{E}_p\Big[\sum_{y'\in\mathscr{Y}}\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{1}\big(Y=y, g(X)=y'\big)\Big]$$

$$= \sum_{y'\in\mathscr{Y}}\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{E}_p\Big[\mathbb{1}\big(Y=y, g(X)=y'\big)\Big] = \sum_{y'\in\mathscr{Y}}\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{E}_p\Big[\mathbb{E}_p\Big[\mathbb{1}\big(Y=y, g(X)=y'\big)\Big|X\Big]\Big]$$

$$= \sum_{y'\in\mathscr{Y}}\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{E}_p\Big[\mathbb{1}\big(g(X)=y'\big)\mathbb{P}\big(Y=y\big|X\big)\Big] = \sum_{y'\in\mathscr{Y}}\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{E}_p\Big[\mathbb{P}\big(Y=y\big|X\big)\mathbb{1}\big(g(X)=y'\big)\Big]$$

$$= \mathbb{E}_p\Big[\sum_{y'\in\mathscr{Y}} \mathbb{1}\big(g(X)=y'\big)\sum_{y\in\mathscr{Y},y\neq y'} \mathbb{P}\big(Y=y\big|X\big)\Big] = \sum_{y'\in\mathscr{Y}} \mathbb{E}_p\Big[\mathbb{1}\big(g(X)=y'\big)\big[1-\mathbb{P}\big(Y=y'\big|X\big)\big]\Big]$$

$$= \sum_{y'\in\mathscr{Y}} \mathbb{P}\big(g(X)=y'\big) - \mathbb{E}_p\Big[\mathbb{1}\big(g(X)=y'\big)\mathbb{P}\big(Y=y'\big|X\big)\Big]$$

$$R(g) = 1 - \textstyle\sum_{y'\in\mathscr{Y}} \mathbb{E}_p\Big[\mathbb{1}\big(g(X)=y'\big)\mathbb{P}\big(Y=y'\big|X\big)\Big]$$

Then $g^*(X) = \arg\min_{g\in F(\mathscr{X},\mathscr{Y})} R(g) = \arg\max_{g\in F(\mathscr{X},\mathscr{Y})} \mathbb{E}_p\Big[\sum_{y\in\mathscr{Y}} \mathbb{1}\big(g(X)=y\big)\mathbb{P}\big(Y=y\big|X\big)\Big]$

Let $P^T = \Big(\mathbb{P}\big(Y=y\big|X\big)\Big)_{y\in\mathscr{Y}}$ and $e_{y'}^T = \Big(\mathbb{1}\big(y'=y\big)\Big)_{y\in\mathscr{Y}}$

$$\sum_{y\in\mathscr{Y}} \mathbb{1}\big(g(X)=y\big)\mathbb{P}\big(Y=y\big|X\big) =< P, e_{g(X)} > \leq \arg\max_{y\in\mathscr{Y}} \mathbb{P}\big(Y=y\big|X\big)$$

Thus $g^*(X) \leq \arg\max_{y\in\mathscr{Y}} \mathbb{P}\big(Y=y\big|X\big)$

Hence $g^*(X) = \arg\max_{y\in\mathscr{Y}} \mathbb{P}\big(Y = y\big|X\big)$

□

5.4 Gaussian naive Bayes classifier

Let $\mathscr{D} = \{(\mathbb{X}_1, Y_1), \cdots, (\mathbb{X}_n, Y_n)\}$ be a given data set, with $\mathbb{X}_i \in \mathbb{R}^d$ and $Y_i \in \mathscr{Y} = \{\lambda_1, \cdots, \lambda_K\}$ classes set for $i \in [\![1, n]\!]$. Let $\mathbb{X}_i = (X_{i1}, \cdots, X_{id})$ with X_{ij} i.i.d for a fixed i and $j \in [\![1, d]\!]$.

We segment our data by class and compute the mean and variance of variable of $\mathbb{X}_i$. Let μ_{kj} and σ^2_{kj} be respectively the mean and the variance of $\mathbb{X}_i$ associated with class $\lambda_k, k \in [\![1, K]\!]$.

For a observation $\mathbb{X} = (X_1, \cdots, X_d)$, X_i i.i.d assume that $X_j = x_j|Y = \lambda_k \sim \mathscr{N}(\mu_{kj}, \sigma_{kj})$ which is $f(x_j|\mu_{kj}, \sigma_{kj}) = \dfrac{1}{\sqrt{2\pi\sigma^2_{kj}}} e^{-\frac{(x_j-\mu_{kj})^2}{2\sigma^2_{kj}}}$

Thus the likelihood is $L(\lambda_k|\mathbb{X}) = \prod_{j=1}^{d} f(x_j|\mu_{kj}, \sigma_{kj})$ and the posterior probability is $\pi(\lambda_k|\mathbb{X}) = \dfrac{L(\lambda_k|\mathbb{X})\pi(\lambda_k)}{\sum_{\lambda'_k\in\mathscr{Y}} L(\lambda'_k|\mathbb{X})\pi(\lambda'_k)}$.

The MAP estimation of λ_k is the gaussian naive Bayes classifier : $\arg\max\limits_{\lambda_k\in\mathscr{Y}} \pi(\lambda_k|\mathbb{X})$.

6 Applications on Poisson processes

6.1 Bayes classifier of "Discrete Poisson process"

For a Poisson process $\{X_t, t \in [0, T]\}$ with rate $\lambda, \lambda > 0$, we discreetize the process in n times as $\{X_t, t \in \{t_1, t_2, \cdots, t_n\}\}$, $X_{t_i}, i \in [\![1, n]\!]$, with $t_{i-1} < t_i$. Hence $X_{t_i} \sim Poisson(\lambda t_i)$.

6.1.1 Bayes classifier of a Poisson distribution

$\mathscr{Y}$ discrete and ordered set

Let $\mathscr{Y} = \{\lambda_1, \cdots, \lambda_K\}$, $p_j = \mathbb{P}(Y = \lambda_j)$ and $N_t^{\lambda_j} \sim Poisson(\lambda_j t)$ which is independent of Y. Assume that $(X_t|Y = \lambda) = N_t^{\lambda}$

$$\mathbb{P}(Y = \lambda_j|X_t = x) = \frac{\mathbb{P}(Y = \lambda_j, X_t = x)}{P(X_t = x)} = \frac{\mathbb{P}(X_t = x|Y = \lambda_j)\mathbb{P}(Y = \lambda_j)}{P(X_t = x)}$$

$$\mathbb{P}(Y=\lambda_j|X_t=x)=\frac{\mathbb{P}(X_t=x|Y=\lambda_j)\mathbb{P}(Y=\lambda_j)}{P(X_t=x)}=\frac{\mathbb{P}(N_t^{\lambda_j}=x)\mathbb{P}(Y=\lambda_j)}{P(X_t=x)}$$

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\max_{\lambda_j}\mathbb{P}(X_t=x|Y=\lambda_j)=\arg\max_{\lambda_j}\mathbb{P}(N_t^{\lambda_j}=x)\mathbb{P}(Y=\lambda_j)$$

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\max_{\lambda_j}\frac{(\lambda_j t)^x}{x!}e^{-\lambda_j t}p_j$$

- if Y is uniformly distributed

  $$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\max_{\lambda_j}\frac{(\lambda_j t)^x}{x!}e^{-\lambda_j t}$$

  Let $f_t(\lambda)=\frac{(\lambda t)^x}{x!}e^{-\lambda t}$

  $$f_t'(\lambda)=e^{-\lambda t}\Big(\frac{(\lambda t)^{x-1}t}{(x-1)!}-\frac{(\lambda t)^x t}{x!}\Big)=e^{-\lambda t}\frac{(\lambda t)^{x-1}t}{(x-1)!}\Big(1-\frac{\lambda t}{x}\Big)$$

λ	0	$\frac{x}{t}$	$+\infty$
f_t	0 ↗	$f_t(\frac{x}{t})$	↘ 0

  If there exists an m s.t $x=\lambda_m t$ then $\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\lambda_m$ and $\lambda_m=\frac{x}{t}$

  Else :

  - If $\forall j, x>\lambda_j t$ Then $\arg\max_{\lambda_j}\mathbb{P}(Y=j|X_t=x)=\arg\max_{\lambda_j}\lambda_j t=\max_j\lambda_j$
  - If $\forall j, x<\lambda_j t$ Then $\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\min_{\lambda_j}\lambda_j t=\min_j\lambda_j$
  - If $\forall j\leq p, x>\lambda_j t, \forall j>p, x<\lambda_j t$ Then $\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\max_{\lambda_j, j\in\{p,p+1\}}f_t(\lambda_j)$

- if Y is not uniformly distributed

  $$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|X_t=x)=\arg\max_{\lambda_j}\frac{(\lambda_j t)^x}{x!}e^{-\lambda_j t}p_j$$

  Which is λ_m such that $\forall j\neq m \qquad \Big(\frac{\lambda_m}{\lambda_j}\Big)^x e^{(\lambda_j-\lambda_m)t}>\frac{p_j}{p_m}$

6.1.2 Applications

For a "discrete Poisson process" in $\{X_{t_i}, i \in [\![1,n]\!]\}$ we can predict a λ for each time in $\{t_1, t_2, \cdots, t_n\}$ and take the rounding value of :

- The mean predicted λ
- The most predicted λ

6.2 Bayes classifier of Poisson process

Proposition 2 : Bayes classifier of Poisson process, $\mathscr{Y}$ discrete

Consider $\{X_t, t \in [0;T]\}$ a Poisson process and $\mathbb{X} = (X_{t_1}, X_{t_2}, ..., X_{t_n-1}, X_T)$ with $0 < t_1 < t_2 < \cdots < t_{n-1} < T$ and $X_{t_i} = x_i$ for $i \in [\![1, n-1]\!]$, $X_T = x_n$.
Let $\mathscr{Y} = \{\lambda_j, j \in [\![1,K]\!]\}$ and $x = (x_1, x_2, \cdots, x_n)$.

Case 1: Y uniformly distributed in $\mathscr{Y}$

The Bayes classifier of the rate λ is $\hat{\lambda} = \dfrac{x_n}{T}$, if there exists an $\lambda_m \in \mathscr{Y}$ such that $x_n = \lambda_m t_n$.
We have $\arg\max\limits_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \lambda_m$ and $\lambda_m = \dfrac{x_n}{T}$.

Otherwise :

- If $\forall \lambda_j \in \mathscr{Y}, x_n > \lambda_j T$ Then $\arg\max\limits_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max\limits_{\lambda_j \in \mathscr{Y}} \lambda_j T = \max\limits_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y}, x_n < \lambda_j T$ Then $\arg\max\limits_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\min\limits_{\lambda_j \in \mathscr{Y}} \lambda_j T = \min\limits_{\lambda_j \in \mathscr{Y}} \lambda_j$
- If $\forall \lambda_j \in \mathscr{Y}, \lambda_j \leq \lambda_m, x_n > \lambda_j T$ and $\forall \lambda_j > \lambda_m, x_n < \lambda_j T$ Then

  $\arg\max\limits_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg \max\limits_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^{x_n} e^{-\lambda_j T}$

Case 2: Y not uniformly distributed in $\mathscr{Y}$

Let $\mathbb{P}(Y = \lambda_j) = p_j$, the Bayes classifier of the rate λ is $\arg\max\limits_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x)$

Which is λ_m such that $\forall j \neq m \qquad \left(\dfrac{\lambda_m}{\lambda_j}\right)^{x_n} e^{(\lambda_j - \lambda_m)T} > \dfrac{p_j}{p_m}$

Proof 1: Since a Poisson process is represented by the values of jump times then we can take $\mathbb{X} = (X_{t_1}, \cdots, X_{t_{n-1}}, X_T)$ with $t_i, i \in [\![1, n-1]\!]$ the jump times (we can also assume $\{t_i, i \in [\![1, n-1]\!]\}$ is any point process in $[0;T]$) .

Let $x = (x_1, \cdots, x_n)$

$\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\mathbb{P}(X_{t_1}=x_1,\cdots,X_{t_{n-1}}=x_{n-1},X_T=x_n|Y=\lambda_j)\mathbb{P}(Y=\lambda_j)$

$\mathbb{P}(X_{t_1}=x_1,\cdots,X_{t_{n-1}}=x_{n-1},X_T=x_n|Y=\lambda_j)$

$=\mathbb{P}(N_{t_1}^{\lambda_j}=x_1,N_{t_2}^{\lambda_j}=x_2,\cdots,N_{t_{n-1}}^{\lambda_j}=x_{n-1},N_T^{\lambda_j}=x_n)$

$=\mathbb{P}(N_{t_1}^{\lambda_j}=x_1,N_{t_2}^{\lambda_j}-N_{t_1}^{\lambda_j}=x_2-x_1,\cdots,N_T^{\lambda_j}-N_{t_{n-1}}^{\lambda_j}=x_n-x_{n-1})$

Because if we know $N_{t_1}^{\lambda_j}=x_1$ we will have the others $N_{t_i}^{\lambda_j}=x_i, i\in[\![2;n]\!]$

$=\prod_{i=1}^{n}\mathbb{P}(N_{t_i}^{\lambda_j}-N_{t_{i-1}}^{\lambda_j}=x_i-x_{i-1})$ ($(N_{t_i}^{\lambda_j}-N_{t_{i-1}}^{\lambda_j})_i$ are i.i.d with $N_{t_0}=0$ and $x_0=0$)

$=\prod_{i=1}^{n}\frac{(\lambda_j(t_i-t_{i-1}))^{x_i-x_{i-1}}}{(x_i-x_{i-1})!}e^{-\lambda_j(t_i-t_{i-1})}$

$=\lambda_j^{\sum_{i=1}^{n}(x_i-x_{i-1})}e^{-\lambda_j\sum_{i=1}^{n}(t_i-t_{i-1})}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{(x_i-x_{i-1})}}{(x_i-x_{i-1})!}$

$\mathbb{P}(X_{t_1}=x_1,\cdots,X_{t_{n-1}}=x_{n-1},X_T=x_n|Y=\lambda_j)=\lambda_j^{x_n}e^{-\lambda_j T}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{(x_i-x_{i-1})}}{(x_i-x_{i-1})!}$

Let $h_X(\lambda)=\lambda^{x_n}e^{-\lambda T}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{(x_i-x_{i-1})}}{(x_i-x_{i-1})!}$

Note that $Y|\mathbb{X}=x$ is discrete .

Compute the derivative of h

$h'_X(\lambda)=\left(x_n\lambda^{x_n-1}e^{-\lambda T}-T\lambda^{x_n}e^{-\lambda T}\right)\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{(x_i-x_{i-1})}}{(x_i-x_{i-1})!}$

$h'_X(\lambda)=\left(x_n-T\lambda\right)\lambda^{x_n-1}e^{-\lambda T}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{(x_i-x_{i-1})}}{(x_i-x_{i-1})!}$

λ	0	$\frac{x_n}{T}$	$+\infty$
h_X	0 ↗	$h_X(\frac{x_n}{T})$	↘ 0

1. Y uniformly distributed in $\mathscr{Y}$

   If there exists an m such that $x_n = \lambda_m T$ then $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \lambda_m$ and $\lambda_m = \dfrac{x_n}{T}$

   Otherwise :

   - If $\forall \lambda_j \in \mathscr{Y}, x_n > \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} h_X(\lambda_j) = \arg\max_{\lambda_j \in \mathscr{Y}} \lambda_j T = \max_{\lambda_j \in \mathscr{Y}} \lambda_j$

   - If $\forall \lambda_j \in \mathscr{Y}, x_n < \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} h_X(\lambda_j) = \arg\min_{\lambda_j \in \mathscr{Y}} \lambda_j T = \min_{\lambda_j \in \mathscr{Y}} \lambda_j$

   - If $\forall \lambda_j \in \mathscr{Y} \leq \lambda_m, x_n > \lambda_j T$ and $\forall \lambda_j > \lambda_m, x_n < \lambda_j T$ Then

     $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^{x_n} e^{-\lambda_j T}$

2. Y not uniformly distributed in $\mathscr{Y}$

   $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j) h_X(\lambda_j)$

   Which is λ_m such that $\forall j \neq m \qquad p_m h_X(\lambda_m) > p_j h_X(\lambda_j)$

   $\iff \dfrac{h_X(\lambda_m)}{h_X(\lambda_j)} > \dfrac{p_j}{p_m} \iff \left(\dfrac{\lambda_m}{\lambda_j}\right)^{x_n} e^{(\lambda_j - \lambda_m)T} > \dfrac{p_j}{p_m}$

□

Proof 2: We consider $\mathbb{X} = \{X_{t_1}, \cdots, X_{t_{n-1}}, X_{t_n}, X_T\}, T_i$ an gamma distribution of the i^{th} jump. Hence $X_{T_i} = i$. Let $x = (1, \cdots, n-1, n, n)$ thus

$\mathbb{P}(Y = \lambda_j | \mathbb{X} = x) = \mathbb{P}(X_{t_1} = 1, \cdots, X_{t_{n-1}} = n-1, X_{t_n} = n, X_T = n | Y = \lambda_j) \mathbb{P}(Y = \lambda_j)$

$= \mathbb{P}(t_1 \in [T_1; T_2[, \cdots t_n \in [T_n, T], N_T^{\lambda_j} = n) \mathbb{P}(Y = \lambda_j)$

$= \mathbb{P}\big((T_1, \cdots, T_n) \in A, N_T^{\lambda_j} = n | Y = \lambda_j\big) \mathbb{P}(Y = \lambda_j)$

Where $A =]0; t_1] \times]t_1; t_2] \times \cdots \times]t_{n-1}; t_n]$

Since

$\mathbb{P}\big((T_1, \cdots, T_n) \in A, N_T^{\lambda_j} = n | Y = \lambda_j\big) = \displaystyle\int_A e^{-\lambda_j T} \lambda_j^n \mathbb{1}_{0 < t_1 < t_2 < \cdots < t_n \leq T} dt_1 dt_2 \cdots dt_n$

By (3) in proposition 1.

$$\mathbb{P}\big((T_1,\cdots,T_n)\in A, N_T^{\lambda_j}=n|Y=\lambda_j\big)=e^{-\lambda_j T}\lambda_j^n\int_A \mathbb{1}_{0<t_1<t_2<\cdots<t_n\leq T}dt_1dt_2\cdots dt_n$$

$$\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\mathbb{P}(Y=\lambda_j)e^{-\lambda_j T}\lambda_j^n\int_A \mathbb{1}_{0<t_1<t_2<\cdots<t_n\leq T}dt_1dt_2\cdots dt_n$$

Note that $Y|((T_1,\cdots,T_n)\in A, N_T^{\lambda_j}=n)$ is discrete.

Since we have $0<t_1<t_2<\cdots<t_n\leq T$:

$$\int_A dt_1dt_2\cdots dt_n=\prod_{i=1}^{n}(t_i-t_{i-1})$$

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|\mathbb{X}=x)=\arg\max_{\lambda_j}e^{-\lambda_j T}\lambda_j^n\mathbb{P}(Y=\lambda_j)\prod_{i=1}^{n}(t_i-t_{i-1})$$

with $t_0=0$ and we will have the same result with $h_X(\lambda_j)=e^{-\lambda_j T}\lambda_j^n\prod_{i=1}^{n}(t_i-t_{i-1})$ □

6.3 Maximum Likelihood Estimation of the rate λ

6.3.1 Maximum Likelihood Estimation MLE

Theorem: The maximum likelihood estimation of the rate λ for a Poisson process $\{X_t, t\in[0;T]\}$ using the distribution of the inter-arrival times I_i or the arrival times T_i is $\hat{\lambda}=\dfrac{N_T}{T}$ with N_T the value of the process at time T.

Proof 1: Using the distribution of the inter-arrival times I_i

Since $I_j\sim exp(\lambda)$ i.i.d and the density of $(I_1,\cdots,I_n,N_T=n)$ is $f_{(I_1,\cdots,I_n,N_T=n)}(i_1,\cdots,i_n,N_T=n)=\lambda^n e^{-\lambda T}\mathbb{1}_{(t_n\leq T)}$ by (2) in proposition 1. Hence his likelihood is

$$L(\lambda|I_1,\cdots,I_n,N_T=n)=\lambda^n e^{-\lambda T}\mathbb{1}_{(t_n\leq T)}$$

We assume that $t_n\leq T$, it's our case then

$$L'(\lambda|I_1,\cdots,I_n,N_T=n)=\lambda^{n-1}\big(n-\lambda T\big)e^{-\lambda T}$$

$$L'(\lambda|I_1,\cdots,I_n,N_T=n)=0\iff\lambda=\frac{n}{T}$$

$$\hat{\lambda}=\arg\max_{\lambda}L(\lambda|I_1,\cdots,I_n,N_T=n)=\frac{n}{T}$$

Since $N_T=n$ then $\hat{\lambda}=\dfrac{N_T}{T}$ □

Proof 2: Using the distribution of the arrival times T_i

Since $T_i \sim Gamma(i,\lambda)$, $f_{T_i}(t_i) = \dfrac{\lambda e^{-\lambda t_i}(\lambda t_i)^{i-1}}{(i-1)!}$, $t_i \geq 0$, since the density of $(T_1,\cdots,T_n,N_T=n)$ is $f_{(T_1,\cdots,T_n,N_T=n)}(t_1,\cdots,t_n) = \lambda^n e^{-\lambda T}\mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$ by (3) then his likelihood is:

$L(\lambda|T_1,\cdots,T_n,N_T=n) = f_{(T_1,\cdots,T_n,N_T=n)}(t_1,\cdots,t_n) = \lambda^n e^{-\lambda T}\mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$

If we have $0<t_1<\cdots<t_n\leq T$, our case:

$L'(\lambda|T_1,\cdots,T_n,N_T=n) = \lambda^{n-1}(n-\lambda T)e^{-\lambda T}$

$L'(\lambda|T_1,\cdots,T_n,N_T=n) = 0 \iff \lambda = \dfrac{n}{T}$

Since $N_T = n$, then $\hat{\lambda} = \dfrac{N_T}{T}$ □

Proposition 3: $\dfrac{N(t)}{t} \to \lambda$ almost surely when $t \to +\infty$.

Proof: We have $\mathbb{E}[\frac{N(t)}{t}] = \frac{\lambda t}{t} = \lambda$. □

6.3.2 Maximum Likelihood Estimation for classification of λ

Let $\mathscr{Y} = \{\lambda_1,\cdots,\lambda_K\}$, instead computing MLE in $\mathbb{R}$ we compute it in $\mathscr{Y}$ to have the classifier which maximizes the likelihood. We have with the likelihood of arrival or inter-arrival times :

$\arg\max_{\lambda_j\in\mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j\in\mathscr{Y}} h(\lambda_j)$ with $h(\lambda) = \lambda^n e^{-\lambda T}$ and $N_T = n$

If there exists an m such that $n = \lambda_m T$ then $\arg\max_{\lambda_j\in\mathscr{Y}} L(\lambda_j) = \lambda_m$ and $\lambda_m = \dfrac{n}{T}$

Otherwise :

- If $\forall\lambda_j \in \mathscr{Y}, n > \lambda_j T$ Then

  $\arg\max_{\lambda_j\in\mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j\in\mathscr{Y}} h(\lambda_j) = \arg\max_{\lambda_j\in\mathscr{Y}} \lambda_j T = \max_{\lambda_j\in\mathscr{Y}} \lambda_j$

- If $\forall\lambda_j \in \mathscr{Y}, n < \lambda_j T$ Then

  $\arg\max_{\lambda_j\in\mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j\in\mathscr{Y}} h(\lambda_j) = \arg\min_{\lambda_j\in\mathscr{Y}} \lambda_j T = \min_{\lambda_j\in\mathscr{Y}} \lambda_j$

- If $\forall\lambda_j \in \mathscr{Y} \leq \lambda_m, n > \lambda_j T$ and $\forall\lambda_j > \lambda_m, n < \lambda_j T$ Then

  $\arg\max_{\lambda_j\in\mathscr{Y}} L(\lambda_j) = \arg\max_{\lambda_j\in\{\lambda_m,\lambda_{m+1}\}} (\lambda_j)^n e^{-\lambda_j T}$

6.4 Maximum A Posteriori estimation of the rate λ

6.4.1 Maximum A Posteriori with the inter-arrival times I_n

The posterior distribution is $\pi(\lambda|I_1,...,I_n,N_T=n)=\mathbb{P}(Y=\lambda|I_1=i_1,...,I_n=i_n,N_T=n)$.
Let $\mathscr{Y}=\{\lambda_1,\cdots,\lambda_K\}$, the MAP classifier in $\mathscr{Y}$ is

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}\frac{L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)}{\sum_j L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)}$$

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|I_1,\cdots,I_n,N_T=n)\mathbb{P}(Y=\lambda_j)$$

If λ is uniformly distributed it's the MLE classifier :

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|I_1,...,I_n,N_T=n)$$

Otherwise let $\mathbb{P}(Y=\lambda_j)=p_j$:

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|I_1,...,I_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}p_j(\lambda_j)^n e^{-\lambda_j T}$$

Which is λ_m such that $\forall\, j\neq m \qquad \left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j-\lambda_m)t_n}>\frac{p_j}{p_m}$

6.4.2 Maximum A Posteriori with the arrival times T_n

The posterior distribution is $\pi(\lambda|T_1,..,T_n,N_T=n)=\mathbb{P}(Y=\lambda|T_1=t_1,..,T_n=t_n,N_T=n)$.
Let $\mathscr{Y}=\{\lambda_1,\cdots,\lambda_K\}$, the MAP classifier in $\mathscr{Y}$ is

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}\frac{L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)}{\sum_j L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)}$$

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|T_1,...,T_n,N_T=n)\mathbb{P}(Y=\lambda_j)$$

If λ is uniformly distributed it's the MLE classifier :

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}L(\lambda_j|T_1,...,T_n,N_T=n)$$

Otherwise let $\mathbb{P}(Y=\lambda_j)=p_j$

$$\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=\arg\max_{\lambda_j\in\mathscr{Y}}p_j\lambda_j^n e^{-\lambda_j T}\mathbb{1}_{(0<t_1<\cdots<t_n\leq T)}$$

if $0<t_1<\cdots<t_n$, our case: $\arg\max_{\lambda_j\in\mathscr{Y}}\pi(\lambda_j|T_1,...,T_n,N_T=n)=\lambda_m$

such that $\forall\, j\neq m \qquad \left(\frac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j-\lambda_m)t_n}>\frac{p_j}{p_m}$

6.5 Machine Learning classifier: Random forest, Gaussian naive Bayes

Data: For a Poisson process $\{X_t, t \in [0,T]\}$, we discrete the process in n times as $\{X_t, t \in \{t_1, t_2, \cdots, t_n\}\}$, $X_{t_i}, i \in [1,n]$. The times $t_1, t_2, \cdots, t_n$ of the process will be our features and the parameter λ our label. And we simulate several process to make our data base. And make a model based Poisson mixture model.

7 Applications on Poisson point processes

7.1 Bayes classifier of homogeneous Poisson point process

Proposition 4.1: Bayes classifier of homogeneous Poisson point process

Consider $X = \{\xi_1, \cdots, \xi_n\}$, $\xi_i \sim Uniform(S)$, $S \subset \mathbb{R}^d$ a bounded Borel measurable set, which form a homogeneous Poisson point process $N(S) = \#X \sim Poisson(\lambda|S|)$, the number of points of n existing in S and $|\cdot|$ a non-negative measure.

Case 1: $\mathscr{Y}$ discrete and ordered set, λ uniformly distributed

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$ is $\hat{\lambda} = \dfrac{n}{|S|}$, if there exist $\lambda_j \in \mathscr{Y}$ such that $\lambda_j = \dfrac{n}{|S|}$. We have $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \hat{\lambda}$.

Otherwise :

- If $\forall \lambda_j, n > \lambda_j|S|$ Then $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg\max_{\lambda_j} \lambda_j|S| = \max_j \lambda_j$
- If $\forall \lambda_j, n < \lambda_j|S|$ Then $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg\min_{\lambda_j} \lambda_j|S| = \min_j \lambda_j$
- If $\forall\, j \leq m\ n > \lambda_j|S|$ otherwise , $n < \lambda_j|S|$ Then

  $$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S) = n) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} \frac{(\lambda|S|)^n}{n!} e^{-\lambda|S|}$$

Case 2: $\mathscr{Y}$ discrete and ordered set, Y no uniformly distributed

Let $\mathbb{P}(Y = \lambda_j) = p_j$, the Bayes classifier of the rate λ is $\arg\max_{\lambda_j \in \mathscr{Y}} \mathbb{P}(Y = \lambda_j | N(S) = n)$

Which is λ_m such that $\forall\, j \neq m \qquad \left(\dfrac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m)|S|} > \dfrac{p_j}{p_m}$

Proof : See **5.1.1**

Proposition 4.2: Bayes classifier of homogeneous Poisson point process

Consider $X_k = \{\xi_1, \cdots, \xi_{n(k)}\}$, $k \in [\![1;K]\!]$, $\xi_i \sim Uniform(S_k)$, $S_1, ..., S_K \subset \mathbb{R}^d$ a collection of disjoint bounded Borel sets which form a homogeneous Poisson point processes $N(S_k) = \#X_k \sim Poisson(\lambda|S_k|)$. Consider $n = \sum_{k=1}^{K} n(k)$ and $|S| = \sum_{k=1}^{K} |S_k|$.

Case 1: $\mathscr{Y}$ discrete and ordered set, Y uniformly distributed

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$ is $\hat{\lambda} = \dfrac{\sum_{k=1}^{K} n(k)}{\sum_{k=1}^{K} |S_k|} = \dfrac{n}{|S|}$, if there exist $\lambda_j \in \mathscr{Y}$ such that $\lambda_j = \dfrac{n}{|S|}$.

We have $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \hat{\lambda}$.

Otherwise :

- If $\forall\, j, n > \lambda_j |S|$ Then

  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \arg\max_{\lambda_j} \lambda_j |S| = \max_j \lambda_j$

- If $\forall \lambda_j, n < \lambda_j |S|$ Then

  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \arg\min_{\lambda_j} \lambda_j |S| = \min_j \lambda_j$

- If $\forall\, j \leq m, n > \lambda_j |S|$ and otherwise, $n < \lambda_j |S|$ Then

  $\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \arg \max_{\lambda_j \in \{\lambda_m, \lambda_{m+1}\}} (\lambda_j)^n e^{-\lambda_j |S|}$

Case 2: $\mathscr{Y}$ discrete and ordered set, Y no uniformly distributed

Assume that $\mathscr{Y} = \{\lambda_1, \cdots, \lambda_J\}$ and $\mathbb{P}(Y = \lambda_j) = p_j$.

The Bayes classifier of the rate λ in the discrete set $\mathscr{Y}$:

$$\arg\max_{\lambda_j} \mathbb{P}(Y = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \arg\max_{\lambda_j} p_j \lambda_j^n e^{-\lambda_j |S|} \prod_{k=1}^{K} \frac{|S_k|}{n(k)!}$$

Which is λ_m such that $\forall\, j \neq m \qquad \left(\dfrac{\lambda_m}{\lambda_j}\right)^n e^{(\lambda_j - \lambda_m)|S|} > \dfrac{p_j}{p_m}$

Proof :

$\mathbb{P}(Y = \lambda_j | N(S_i) = n(i), i = 1, \ldots, K) = \mathbb{P}(N(S_i) = n(i), i = 1, \ldots, K | Y = \lambda_j) \mathbb{P}(Y = \lambda_j)$
Case 1: Since Y is uniformly distributed then

$\mathbb{P}(Y = \lambda_j | N(S_i) = n(i), i = 1, \ldots, K) = \mathbb{P}(N(S_i) = n(i), i = 1, \ldots, K | Y = \lambda_j)$

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|N(S_1)=n(1),...,N(S_K)=n(K))=\arg\max_{\lambda_j}\prod_{i=1}^{K}\frac{(\lambda|S_i|)^{n_i}}{n_i!}e^{-\lambda|S_i|}$$

$$=\arg\max_{\lambda_j}\lambda_j^n e^{-\lambda_j|S|}\prod_{k=1}^{K}\frac{|S_k|}{n(k)!}=\arg\max_{\lambda_j}(\lambda_j)^n e^{-\lambda_j|S|}$$

Case 2: Since Y is not uniformly distributed then

$$\arg\max_{\lambda_j}\mathbb{P}(Y=\lambda_j|N(S_1)=n(1),...,N(S_K)=n(K))=\arg\max_{\lambda_j}p_j\prod_{i=1}^{K}\frac{(\lambda|S_i|)^{n_i}}{n_i!}e^{-\lambda|S_i|}$$

$$=\arg\max_{\lambda_j}p_j(\lambda_j)^n e^{-\lambda_j|S|}$$

□

7.2 Bayes classifier of inhomogeneous Poisson point process

Proposition 5.1: Bayes classifier of homogeneous Poisson point process

Consider $X=\{\xi_1,\cdots,\xi_n\}$, $\xi_i\sim\frac{\lambda(x_i)}{\Lambda(S)}$, $S\subset\mathbb{R}^d$ a bounded Borel measurable set, which form a homogeneous Poisson point process $N(S)=\#X\sim Poisson(\Lambda(S))$, the number of points of n existing in S.

Let $\mathscr{F}=\{\lambda_1,...,\lambda_J\}$ be a discrete set of measurable function in $\mathbb{R}^d$, uniformly distributed. The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\hat{\lambda}=\arg\max_{\lambda\in\mathscr{F}}\mathbb{P}(F=\lambda|N(S)=n)=\arg\max_{\lambda\in\mathscr{F}}\frac{(\Lambda(S))^n}{n!}e^{-\Lambda(S)}.$$

Proposition 5.2: Bayes classifier of inhomogeneous Poisson point process

Consider $X_k=\{\xi_1,\cdots,\xi_{n(k)}\}$, $k\in[\![1;K]\!]$, $\xi_i\sim\frac{\lambda(x_i^k)}{\Lambda(S_k)}$, $S_1,...,S_K\subset\mathbb{R}^d$ a collection of disjoint bounded Borel sets which form a inhomogeneous Poisson point processes $N(S_k)=\#X_k\sim Poisson(\Lambda(S_k))$. Consider $n=\sum_{k=1}^{K}n(k)$ and $|S|=\sum_{k=1}^{K}|S_k|$.

Case 1: F uniformly distributed

Let $\mathscr{F}=\{\lambda_1,...,\lambda_J\}$ be a discrete set of measurable function in $\mathbb{R}^d$, uniformly distributed. The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\hat{\lambda}=\arg\max_{\lambda\in\mathscr{F}}\mathbb{P}(F=\lambda|N(S_1)=n(1),...,N(S_K)=n(K))=\arg\max_{\lambda\in\mathscr{F}}\prod_{k=1}^{K}\frac{(\Lambda(S_k))^{n(k)}}{n(k)!}e^{-\Lambda(S_k)}\ .$$

Case 2: F no uniformly distributed

Assume that $\mathbb{P}(F=\lambda_j)=p_j$ and $\mathscr{F}$ no uniformly distributed.

The Bayes classifier of the rate λ in the discrete set $\mathscr{F}$ is

$$\hat{\lambda} = \arg\max_{\lambda_j} \mathbb{P}(F = \lambda_j | N(S_1) = n(1), ..., N(S_K) = n(K)) = \arg\max_{\lambda_j} p_j \prod_{k=1}^{K} \frac{(\Lambda_j(S_k))^{n(k)}}{n(k)!} e^{-\Lambda_j(S_k)}$$

.

Which is λ_m such that $\forall\, j \neq m \qquad \prod_{k=1}^{K} \left(\frac{\Lambda_m(S_k)}{\Lambda_j(S_k)}\right)^{n(k)} \exp[(\Lambda_j(S_k) - \Lambda_m(S_k)] > \frac{p_j}{p_m}$

Proposition 5.3 (Lemma) Bayes classifier of an inhomogeneous Poisson process in a separable and bounded metric space

Let $(X, Y) \in S^\infty \times \{0, \ldots, M\}$. Let $X|Y = j$ be Poisson processes on S^∞ with intensities λ_j, $j = 1, \ldots, M$, respectively. Then the Bayes rule classifies a point $x \in S^\infty$ into class j if

$$\forall_{i \neq j} \quad \exp\{\mu_i(S) - \mu_j(S)\} \prod_{\xi \in x} \frac{\lambda_j(\xi)}{\lambda_i(\xi)} > \frac{p_i}{p_j}$$

where $p_i = \Pr(Y = i)$, $i = 1, \ldots, M$ and as before, $\mu_i(S) = \int_S \lambda_i(\xi) d\nu(\xi)$, $i = 1, \ldots, M$.[2]

8 Results

8.1 Analysis of probabilistic models results

We choose λ in $\{1,2,3,4,5,6,7,8,9,10\}$ and N processes for each one .

Table 1: Precisions of Bayes classifier 'discrete' (mean) by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	50.33%	27%	16%	14.33%	14%	14%	17%	16.67%	22%	27.67%	21.9%
5	75%	47%	39%	32.33%	31%	27%	29%	27.33	30.33	41.33	37.93
10	81%	57%	50.33%	40.33%	38.33%	40%	29%	30.33%	36.33%	53.67%	45.63%
15	87.67%	70.67%	59.67%	52.33%	47.33%	46.33%	35%	44.33%	43.67%	52.67%	53.97%
20	89.33%	77.67%	70.67%	62.33%	54.33%	49.33%	53.33%	47.33%	41%	63%	60.83%
25	94%	83%	74.33%	67%	62%	53.33%	51.67%	51.67%	54.67%	62.67%	65.43%
30	95%	86%	79.67%	69.33%	62%	63%	56%	48.67%	55.67%	69.67%	68.5%
50	99.33%	95.33%	86%	85.33%	76.33%	73%	68.33%	67%	73%	74.67%	79.83%
100	100%	99.67%	97%	96%	93.67%	90.67%	88.67%	86.67%	86%	86.33%	92.47%
200	100%	100%	100%	99.67%	97.67%	97%	97%	95.67%	92.67%	94%	97.37%

Table 2: Precisions of Bayes classifier 'discrete' (most) by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	85.67%	4.33%	24.33%	4.33%	11.67%	11.67%	11.33%	0.67%	4.67%	67.67%	22.63%
5	88%	45.33%	47%	34.33%	34.33%	31.33%	31%	22%	21.33%	68%	42.27%
10	88.67%	60.33%	58.33%	47%	41%	45%	37.67%	35%	27.67%	73%	51.37%
15	92%	77%	70%	59.67%	54%	54.33%	43.67%	50.67%	39.67%	73%	61.4%
20	92.67%	80.33%	74.33%	69%	61%	57%	57%	47%	45%	76.67%	66%
25	96.33%	85%	79%	72.67%	67.67%	58%	56.67%	58.67%	52.67%	78%	70.47%
30	97%	87%	83%	74.33%	68.33%	68%	65%	54.67%	55%	80.33%	73.27%
50	99.33%	96.33%	90%	86.33%	81%	77.33%	74%	72%	72.67%	82.33%	83.13%
100	100%	100%	97.33%	97%	94.33%	92.33%	90.33%	88.33%	85.33%	91%	93.6%
200	100%	100%	100%	99.33%	98.33%	97.67%	98.67%	96.67%	92.67%	94.33%	97.77%

Table 3: Precisions of Bayes classifier 'inter-arrival times' by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	42.67%	16.67%	15.67%	18.67%	17.67%	16.33%	11%	14.67%	15%	66.33%	23.47%
5	66.33%	52.67%	48%	42%	34.33%	31%	32%	27.33%	31.67%	60%	42.53%
10	79.33%	68.33%	55.33%	57%	50%	50%	38.33%	39.67%	39.33%	71.33%	54.87%
15	90.67%	81.33%	74.33%	67.33%	61.33%	56.67%	54.33%	51%	47%	73.33%	65.73%
20	96%	82.67%	76%	66%	66%	62.67%	55%	52.67%	54.67%	78%	68.97%
25	96.67%	91.67%	87.33%	81.33%	74%	71%	66.67%	62.33%	59%	81.33%	77.13%
30	98.33%	95.33%	85.67%	82%	80%	72.67%	71.67%	64.33%	60.33%	82.67%	79.3%
45	100%	98%	94.33%	93%	86.33%	82.67%	78%	72%	72.33%	90%	86.67%
50	100%	97.33%	95%	90.33%	91%	83.67%	79.33%	79%	75%	93%	88.37%
55	99.67%	98.67%	96.33%	93%	92.33%	88.67%	84.67%	78.67%	38.33%	98.67%	86.9%
100	100%	100%	99.33%	99%	97.33%	96.33%	97%	92%	89%	95.33%	96.53%
200	100%	100%	99.67%	100%	100%	99.67%	98.67%	99.67%	99%	99.33%	99.6%

Table 4: Precisions of Bayes classifier 'inter times' by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	76.67	00%	20.33%	14.67%	19.33%	19.67%	17%	13%	12.33%	72%	26.5%
5	85%	43.33%	47.33%	41%	36%	27%	31.33%	28.67%	31.67%	62.67%	43.4%
10	91.33%	62.33%	51.33%	55.67%	49.33%	45.67%	37.33%	38.33%	36.33%	73.33%	54.1%
15	94%	76.67%	74.33%	66.33%	58%	55.33%	53.33%	49%	46.33%	71.67%	64.5%
20	97.67%	89.67%	36.67%	32%	33.33%	56.67%	47.33%	48.33%	38.33%	71%	55.1%
25	98%	86.33%	81%	80.67%	72.67%	68%	63.67%	58%	59%	80.33%	74.77%
30	99.33%	90%	84.67%	79.67%	75%	70.33%	72.33%	62%	59.33%	83%	77.57%
45	100%	98.67%	94.33%	93.67%	86.67%	81.33%	78.67%	73.67%	73%	89.33%	86.93%
50	100%	97.67%	94.67%	91%	89.33%	83.67%	78.67%	79%	81%	91%	88.63%
55	100%	98.67%	96.67%	93.33%	92.67%	46.33%	47%	45.33%	49.67%	98.67%	76.83%
100	100%	100%	99.33%	99%	97.33%	96.67%	97.67%	92%	88.33%	94%	96.43%
200	100%	100%	99.67%	100%	100%	99.67%	98.67%	99.67%	99%	99.33%	99.6%

Table 5: Precisions of MLE classifier (round()) by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	72.67%	26.33%	21.67%	22.33%	15.33%	14.33%	16%	12.67%	14.33%	56.67%	27.23%
5	89%	57.67%	51%	40.33%	40%	33.33%	33.67%	30.33%	30.67%	60.67%	46.67%
10	90.67%	78.67%	54%	60.33%	40%	54.33%	43%	49.33%	35.66%	69.67%	57.57%
15	95.67%	83.67%	76.67%	67%	62%	59.67%	48%	48%	48%	49.67%	70.67%
20	98.67%	88.33%	76.67%	74%	60.33%	64.33%	55.33%	61%	48.33%	73.33	70.03
25	98.67%	92%	83.33%	78%	76.67%	64.67%	64.67%	64%	63.33%	77.33%	76.27%
30	99.67%	95.33%	86%	81.67%	77.67%	76.67%	67.67%	67%	61.67%	84%	79.73%
50	100%	99.67%	94.67%	91.33%	89.67%	87%	79.33%	81%	76.67%	86.67%	88.6%
55	100%	98%	95%	93.67%	86.67%	89.33%	79.67%	83.67%	45.67%	99.67%	87.13%
100	100%	100%	100%	98.67%	97.67%	95.67%	97%	92.67%	91%	92.67%	96.53%
200	100%	100%	100%	100%	99.67%	99.67%	99.67%	99.33%	97.67%	97.67%	99.37%

Table 6: Precisions of MLE (round()) by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	30.33%	26.33%	21.67%	22.33%	15.33%	14.33%	16%	12.67%	14.33%	12.67%	18.6%
5	77.33%	57.67%	51%	40.33%	40%	33.33%	33.67%	30.33%	30.67%	24.33%	41.87%
10	82.67%	78.67%	54%	60.33%	40%	54.33%	43%	49.33%	35.66%	39.67%	53.77%
15	93.33%	83.67%	76.67%	67%	62%	59.67%	48%	48%	48%	47.33%	58.57 %
20	98.67%	88.33%	76.67%	74%	60.33%	64.33%	55.33%	61%	48.33	51.33%	67.7%
25	98%	92%	83.33%	78%	76.67%	64.67%	64.67%	64%	63.33%	55.67%	74.03%
30	99.67%	95.33%	86%	81.67%	77.67%	76.67%	67.67%	67%	61.67%	61.67%	77.5%
50	100%	99.67%	94.67%	91.33%	89.67%	87%	79.33%	81%	76.67%	75.33%	87.47%
100	100%	100%	100%	98.67%	97.67%	95.67%	97%	92.67%	91%	90.67%	96.33%
200	100%	100%	100%	100%	99.67%	99.67%	99.67%	99.33%	97.67%	96.33%	99.23%

8.2 Analysis of ML models results

For the random forest classifier we have 47%,100%,100%,100%,... accuracy for training data.

For the gaussian naive Bayes classifier we have 26%,41.86%,52.19%,60.19%,66.05%,70.90%,74.9%,

84.57%,94.47% and 98.14% accuracy for training data.

Table 7: Precisions of Random forest classifier in test data by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	46.67%	18.89%	18.89%	7.78%	13.33%	14.44%	17.78%	22.22%	23.33%	42.22%	22.55%
5	86.56%	52.22%	41.11%	31.11%	33.33%	26.67%	31.11%	32.22%	31.11%	53.33%	41.78%
10	90%	71.11%	48.89%	65.56%	34.44%	37.78%	42.22%	42.22%	50%	64.44%	54.67%
15	95.56%	90%	73.33%	56.67%	50%	48.89%	50%	56.67%	52.22%	57.78%	63.11%
20	94.44%	86.67%	82.22%	61.11%	68.89%	68.89%	67.78%	41.11%	48.89%	66.67%	68.67
25	96.67%	88.89%	84.44%	68.89%	71.11%	66.67%	64.44%	62.22%	60%	77.78%	74.11%
30	97.78%	95.56%	91.11%	81.11%	70%	76.67%	65.56%	61.11%	57.78%	82.22%	77.89%
50	100%	95.56%	91.11%	91.11%	95.56%	78.89%	83.33%	77.78%	78.89%	84.44%	87.67%
100	100%	100%	100%	98.89%	94.44%	97.78%	96.67%	91.11%	88.89%	95.56%	96.33%
200	100%	100%	100%	100%	100%	100%	98.89%	98.89%	98.89%	97.78%	99.44%

Table 8: Precisions of gaussian naive Bayes classifier in test data by λ and T.

T ↓ / λ →	1	2	3	4	5	6	7	8	9	10	Total
1	75.55%	2%	32.22%	13.33%	11.11%	2%	0%	7.77%	18.89%	54.44%	25%
5	85.56%	41.11%	38.89%	30%	41.11%	27.78%	27.78%	27.78%	23.33%	53.33%	40%
10	90%	65.56%	57.78%	52.22%	40%	45.55%	33.33%	43.33%	31.33%	68.89%	53%
15	93.33%	75.56%	68.89%	63.33%	55.56%	46.67%	52.22%	52.22%	36.67%	58.89%	60%
20	91.11%	86.67%	82.22%	65.56%	63.33%	60%	63.33%	45.56%	46.67%	75.56%	68%
25	94.44%	93.33%	76.67%	75.56%	72.22%	64.44%	53.33%	56.67%	55.56%	76.67%	72%
30	96.67%	90%	85.56%	81.11%	62.22%	64.44%	67.78%	60%	57.78%	77.78%	74%
50	100%	92.22%	90%	87.78%	84.44%	71.11%	84.44%	73.33%	75.56%	77.76%	84.57%
100	100%	100%	100%	96.67%	96.67%	94.44%	91.11%	91.11%	86.67%	95.56%	98.14%
200	100%	100%	100%	98.89%	98.89%	98.89%	98.89%	95.56%	96.67%	93.33%	98.14%

8.3 Inhomogeneous Poisson process in a separable and bounded metric space

Application: We define three gaussian intensities on the space S:=[0, 1] x [0, 1].

Higher intensity around (0.25, 0.25) :$\lambda_1(x,y) = 10e^{-\frac{(x-0.25)^2+(y-0.25)^2}{2\times 0.05^2}} + 0.5$

Higher intensity around (0.75, 0.75): $\lambda_2(x,y) = 8e^{-\frac{(x-0.75)^2+(y-0.75)^2}{2\times 0.08^2}} + 0.2$

Higher intensity around (0.25, 0.75) and (0.75, 0.25):

$\lambda_3(x,y) = 7e^{-\frac{(x-0.25)^2+(y-0.25)^2}{2\times 0.06^2}} + 7e^{-\frac{(x-0.75)^2+(y-0.75)^2}{2\times 0.06^2}} + 1$

And we compare the results of classifications Bayes classifier using the intensities to the KNN classifier.

9 Discussion

The results show us that the models have the same evolution following T. The precisions increase with T. If T is less than 10 the precisions are under 50% , at $T = 50$ we reach mostly 85% and at $T = 100$ we reach mostly 96%. The Bayes classifier for Poisson process and the MLE classifier reach 99% at $T = 200$. And we know that the MLE converge to the true λ when T tend to infinity.

For the spatial point Poisson classification the KNN classifier has better accuracy but it doesn't give us a density function. And Bayes classifier can have a good accuracy and

give a density function which permit to do simulation.

10 Conclusion

We conclude that for Poisson processes the Bayes classifier is better, and it's the right solution. However the others classifiers are good too,we have almost the same score especially with the MLE classifier. We know that the MLE converge to λ when T tend to infinity.

And for spatial point Poisson processes, if we want to just classify the ML classifier is better. But to have better simulation we have to use the Bayes classifier of densities.

References

[1] Advanced Data Sciences course - Mathematics and Computer science Master 2 track, Gustave Eiffel University, Mohamed Hebiri

[2]Alejandro Cholaquidis and al, On the classification problem for Poisson point processes

[3] Christian P. Robert, Le choix bayesien principes et pratiques, Collection Statistique et probabilités appliquées, Springer

[4] Cottrel-Duhamel, Exercices de probabilités

[5] Esra Bas, Basics of Probability and Stochastic Processes, Springer Nature Switzerland AG 2019

[6] Poisson point process https://en.wikipedia.org/wiki/Poisson_point_process

[7] Bayesian statistics https://en.wikipedia.org/wiki/Bayesian_statistics

[8] Naive Bayes classifier https://en.wikipedia.org/wiki/Naive_Bayes_classifier

11 Appendices

11.1 Abbreviation

i.i.d : indepenpendent and identically distributed

MCF : Maître de conférence

Pr: Professor

11.2 Github Link

https://github.com/DIOUFY7/Classification-Poisson-processes-and-spatial-Poisson-pro

11.3 Random forest classifier

https://en.wikipedia.org/wiki/Random_forest

11.4 K-Nearest Neighbors KNN classifier

https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

11.5 Confusion matrix of Poisson processes classification models

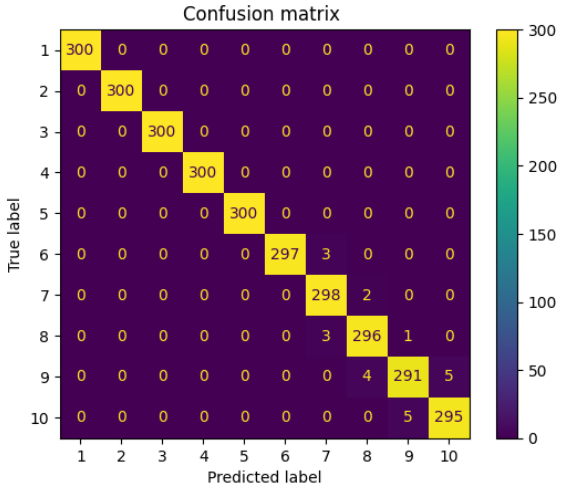


Figure 1: inter times with T=200

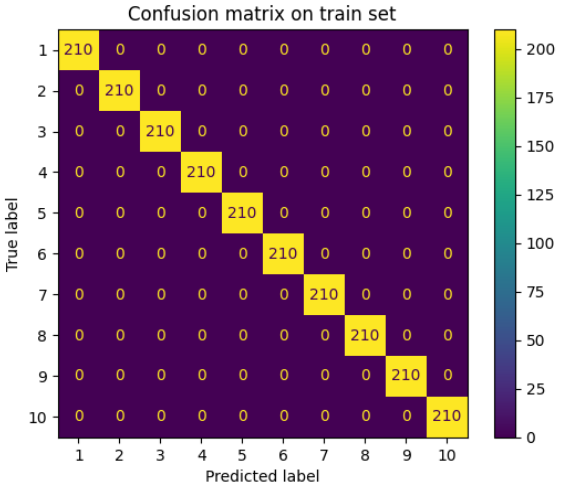


Figure 2: random forest on train set with T=200

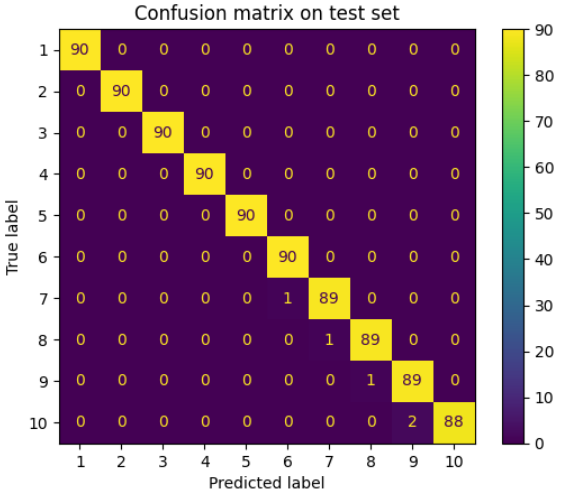


Figure 3: random forest on test set with T=200

11.6 Confusion matrix of spatial Poisson point processes classification models

Results and K neighbors KNN

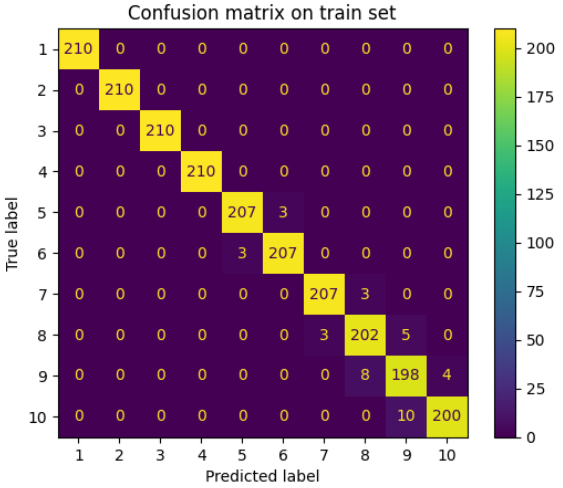


Figure 4: Gaussian naive Bayes on train set with T=200

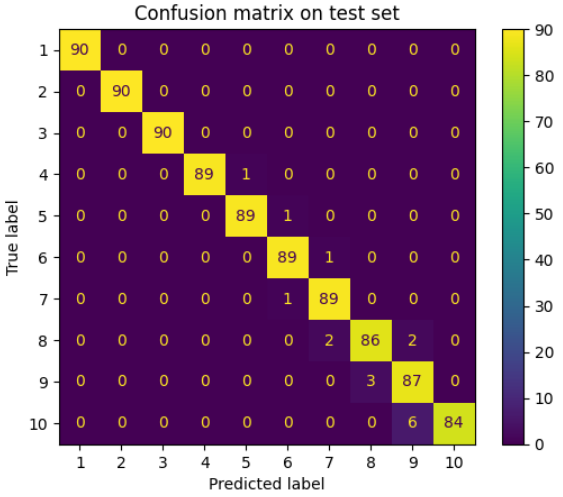


Figure 5: Gaussian naive Bayes on test set with T=200

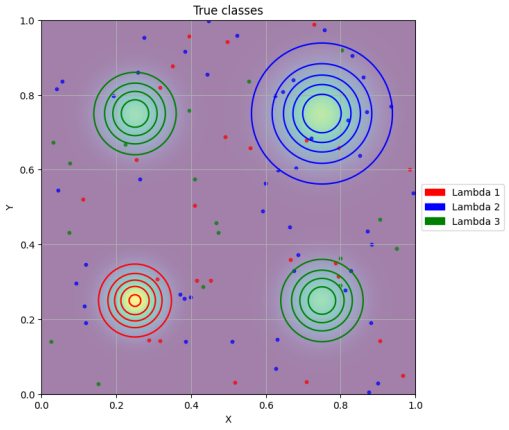


Figure 6: Spatial point processes

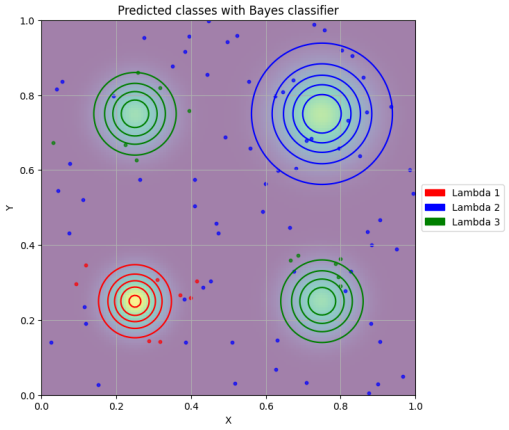


Figure 7: Spatial point processes predicted class Bayes

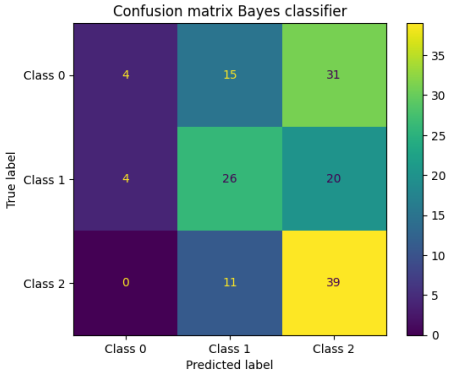


Figure 8: Confusion matrix Spatial point

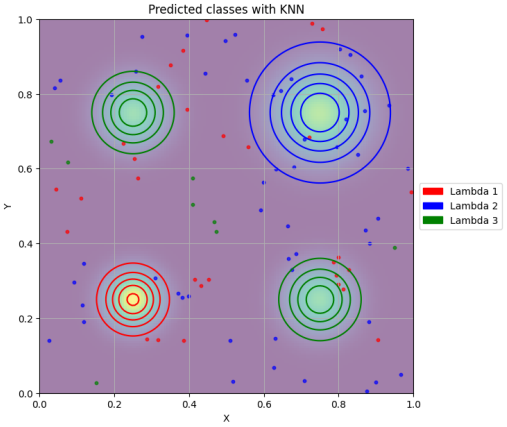


Figure 9: KNN classifications

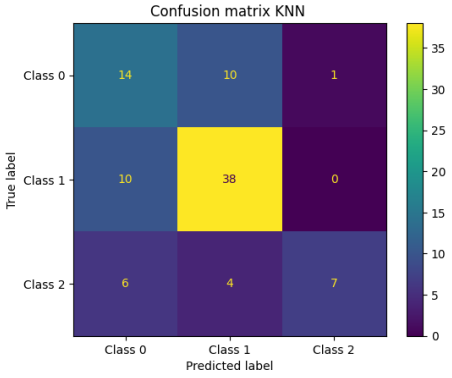


Figure 10: Confusion matrix KNN