On the classification problem for Poisson Point Processes.

Alejandro Cholaquidis

CMAT-Facultad de Ciencias, UdelaR

Escuela CIMPA - La Habana

1. Binary Classification
   - On the classification problem

2. Classification of P.P.P.
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3. Simulations

1. Binary Classification
   - On the classification problem

2. Classification of P.P.P.
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3. Simulations

On the classification problem

Goal:

From a training sample $\mathcal{D}_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ i.i.d. of $(X, Y) \in \mathcal{F} \times \{0, 1\}$ we want a predictor $g : \mathcal{F} \to \{0, 1\}$, which *minimize* $\mathbb{P}\big(g(X) \neq Y\big)$.

Let us denote

1) $\eta(x) = \mathbb{E}(Y|X = x) = \mathbb{P}(Y = 1|X = x)$ the regression function.
2) $g^*(x) = \mathbb{I}_{\{\eta(x)>1/2\}}$ the Bayes rule.
3) $L^* = \mathbb{P}\big(g^*(X) \neq Y\big)$ the optimal Bayes risk.

If $\eta_n(x) : \mathcal{E} \to [0, 1]$ and $g_n(x) = \mathbb{I}_{\{\eta_n(x)>1/2\}}$

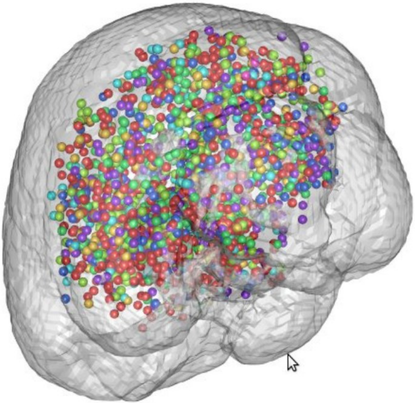
$$0 \leq \mathbb{P}(g_n(X) \neq Y) - L^* \leq \mathbb{E}|\eta(X) - \eta_n(X)|^2.$$

The classical estimator of $\eta(x)$ is $\eta_n(X) = \sum_{i=1}^n W_{ni}(X)Y_i$. for some weights $W_{ni}(X) = W_{ni}(X, X_1, \ldots, X_n)$.

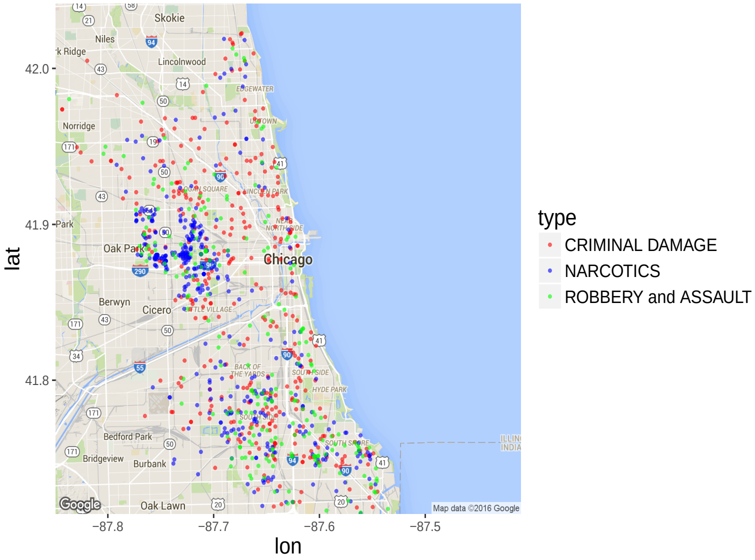
k-NN in $\mathbb{R}^d$

For $W_{ni}(X) = \frac{1}{k}\mathbb{I}_{\{X_i \in k_n(X)\}}$, where $X_i \in k_n(X)$ if X_i is one of the k nearest neighbours of X, conditions 1 to 5 holds if $n \to \infty$ and $k_n/n \to 0$.

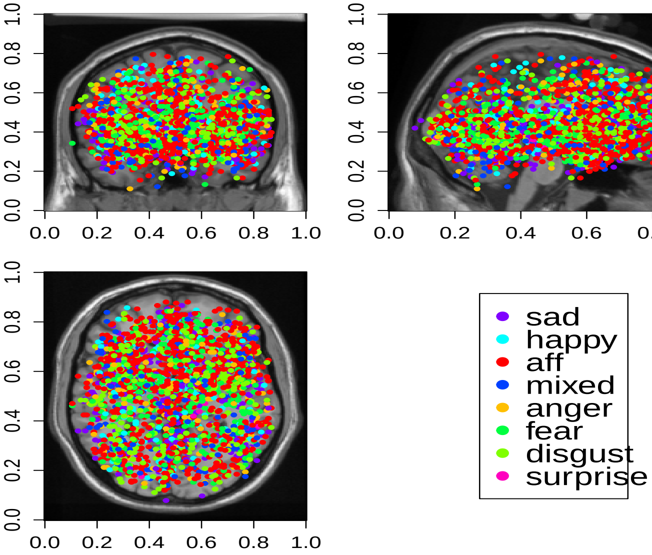
Motivation



Motivation


Skokie
Niles
Lincolnwood
EDGEWATER
Norridge
UPTOWN
LOGAN SQUARE
LINCOLN PARK
NEAR NORTH SIDE
Oak Park
Chicago
Berwyn
Cicero
LITTLE VILLAGE
BACK OF THE YARDS
SOUTH SIDE
HYDE PARK
Bedford Park
Bridgeview
Burbank
SOUTH SHORE
Oak Lawn
Map data ©2016 Google
lat
lon
42.0
41.9
41.8
-87.8
-87.7
-87.6
-87.5
type
CRIMINAL DAMAGE
NARCOTICS
ROBBERY and ASSAULT

Motivation


0.0
0.2
0.4
0.6
0.8
1.0
sad
happy
aff
mixed
anger
fear
disgust
surprise

P.P.P.

Given,

- (S, ρ) a separable bounded metric space.
- ν a Borel measure.
- $S^\infty = \{x \subset S : \#x < \infty\}$.
- $\lambda : S \to \mathbb{R}^+$ integrable.
- $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space.

$X : \Omega \to S^\infty$ is a Poisson Point Process on S with intensity λ, $X \sim \mathit{Poisson}(S, \lambda)$, if

- $N_A : \Omega \to \{0, 1, \ldots, \infty\}$, $N_A(\omega) = \#(X(\omega) \cap A)$ are random variables for all $A \in \mathcal{B}(S)$.
- given n disjoint Borel subsets $A_1, \ldots, A_n$ of S, $N_{A_1}, \ldots, N_{A_n}$ are independent.
- N_A has Poisson distribution with mean $\mu(A)$, being

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

1. Binary Classification
   - On the classification problem

2. Classification of P.P.P.
   - Bayes rule
   - k-NN
   - Other distances for k-NN

3. Simulations

Radon Nikodym for P.P.P.

Theorem

Let $X_1 \sim Poisson(S, \lambda_1)$ and $X_2 \sim Poisson(S, \lambda_2)$ being (S, ρ) a non-empty, bounded metric space. such that $\mu_i(S) < \infty$, $i = 1, 2$. Suppose that $\lambda_1(\xi) > 0 \Rightarrow \lambda_2(\xi) > 0$. Then, $P_{X_1} \ll P_{X_2}$, with density

$$f(x) = \exp\left[\mu_2(S) - \mu_1(S)\right] \prod_{\xi \in x} \frac{\lambda_1(\xi)}{\lambda_2(\xi)},$$

with $0/0 = 0$.

Corollary

If $X_2 \sim Poisson(S, 1)$, then, for all $X \sim Poisson(S, \lambda)$ we have $P_X \ll P_{X_2}$ and

$$f(x) = \exp\left[\nu(S) - \mu(S)\right] \prod_{\xi \in x} \lambda(\xi),$$

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

Bayes rule

$$\mathbb{P}(Y=1|X=x) = \frac{f_{X_1}(x)\mathbb{P}(Y=1)}{f_{X_1}(x)\mathbb{P}(Y=1)+f_{X_0}(x)\mathbb{P}(Y=0)},$$

then,

$$\mathbb{P}(Y=1|X=x) > \frac{1}{2} \quad \Leftrightarrow \quad \exp\left[\mu_0(S)-\mu_1(S)\right]\prod_{\xi\in x}\frac{\lambda_1(\xi)}{\lambda_0(\xi)} > \frac{(1-p)}{p},$$

where $p=\mathbb{P}(Y=1)$.

For two homogeneous Poisson processes

$$\exp\left((\lambda_0-\lambda_1)\nu(S)\right)\left(\frac{\lambda_1}{\lambda_0}\right)^n > \frac{1-p}{p}, \tag{1}$$

where $n=\#x$.

Bayes rule: estimating the intensity

Given a realization $\{\xi_1, \ldots, \xi_n\}$ of X, the intensity $\lambda(x)$ for $x \in S$ can be estimated by

$$\hat{\lambda}(x) = \frac{1}{K_\sigma(x)} \sum_{i=1}^{n} k\left(\frac{\rho(x, \xi_i)}{\sigma}\right) \qquad K_\sigma(x) = \int_S k\left(\frac{\rho(x, \xi)}{\sigma}\right) d\nu(\xi),$$

with $k : \mathbb{R}^d \to \mathbb{R}$ a symmetric kernel, $\sigma > 0$ a smoothing parameter.

If we have $\{X_1, \ldots, X_m\}$, where $X_j = \{\xi_1, \ldots, \xi_{n(j)}\}$ $j = 1, \ldots, m$, let us define $k_{\sigma_m}(\cdot) = k(\cdot/\sigma_m)$,

$$\hat{\hat{\lambda}}_m(x) = \frac{1}{m} \sum_{j=1}^{m} \hat{\lambda}_j^m(x) \qquad \hat{\lambda}_j^m(x) = \frac{1}{K_{\sigma_m}(x)} \sum_{i=1}^{n(j)} k_{\sigma_m}\big(\rho(x, \xi_i)\big).$$

Theorem

If $\sigma_m \to 0$, we have that $\lim_{m\to\infty} \sup_{x\in S} |\hat{\hat{\lambda}}_m(x) - \lambda(x)| = 0$ *a.s.*

A suitable distance for k-NN

Hausdorff distance.

Given two non-empty compact sets $A, C \subset S$,

$$d_H(A, C) = \max\left\{ \sup_{a\in A} d(a, C),\ \sup_{c\in C} d(c, A) \right\}, \tag{2}$$

where $d(a, C) = \inf\{\rho(a, c) : c \in C\}$.
Remark: (S^∞, d_H) is separable.

1 Binary Classification
- On the classification problem

2 Classification of P.P.P.
- Bayes rule
- k-NN
- Other distances for k-NN

3 Simulations

Theorem, Consistency of k-NN

Let us consider

- $(X, Y) \in S^{\infty} \times \{0, 1\}$.
- $X_1 \doteq X|Y = 1 \sim Poisson(\lambda_1), \quad X_0 \doteq X|Y = 0 \sim Poisson(\lambda_0)$.
- λ_1, and λ_0 continuous functions of ρ.
- ν does not have atoms (i.e. $\nu(\{z\}) = 0$ for all $z \in S$).

Then $k - NN$ is consistent.

Simultations

The model

We generate $N = 200$ data ($N/2$ for training), half of them (the 0-class) with Poisson distribution on $[0, 1]^2$ and intensity

$$\lambda_0(x, y) = c_0 \exp(-d_0((x - 1/2)^2 + (y - 1/2)^2)),$$

where we fixed $c_0 = 500$ and $d_0 = 20$.

$$\lambda_1((x, y), c_1, d_1, a) = c_1 \exp(-d_1((x - 1/2 - a_1)^2 + (y - 1/2 - a_2)^2)).$$

Level sets of the estimation of the densities

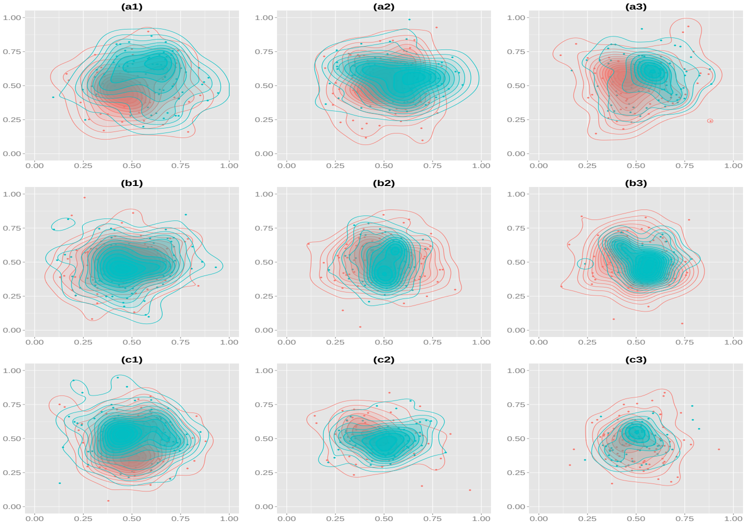


Figure: In all cases $d_0 = 20$, $c_0 = 500$. **First row:** $x = 0.03$, $x = 0.04$, $x = 0.05$. $d_1 = 20$, $c_1 = 500$. **Second row:** $x = 0$, $d_1 = 20, 30, 40$, $c_1 = 600$. **Third row:** $x = 0$, $d_1 = 20, 30, 40$ $c_1 = 700$.