Machine Learning for Graphs and Sequential Data Exercise Sheet 03

Temporal Point Processes

Problem 1: Consider a temporal point process, where all the inter-event times $\tau_i = t_i - t_{i-1}$ are sampled i.i.d. from the distribution with the survival function

$$S(\tau) = \exp\Big(-(e^{b\tau} - 1)\Big)$$

with a parameter $b > 0$.

a) Write down the closed-form expression for the conditional intensity function $\lambda^*(t)$ of this TPP. Simplify as far as you can.

b) Write down the closed-form expression for the log-likelihood of a sequence $\{t_1, ..., t_N\}$ generated from this TPP on the interval $[0, T]$. Simplify as far as you can.

Problem 2: Consider an inhomogeneous Poisson process (IPP) on $[0, 1]$ with the intensity function $\lambda(t) = 2t$. We simulate a sample from this IPP using thinning. For this, we first simulate a *homogeneous* Poisson process (HPP) with intensity $\mu = 4$ and apply the thinning procedure described in the lecture. What is the expected number of events from the HPP that will be rejected when using this procedure?

Problem 3: Consider an inhomogeneous Poisson process on $[0, 4]$ with the intensity function $\lambda(t) = \beta t$, where $\beta > 0$ is a parameter that has to be estimated. You have observed a single sequence $\{1, 2.1, 3.3, 3.8\}$ generated from this IPP. What is the maximum likelihood estimate of the parameter β?

Problem 4: Consider a *neural* temporal point process where the conditional intensity function is defined with a neural network. In particular, for a time point t_i, we represent the history $\{t_1, t_2, \ldots, t_{i-1}\}$ with a fixed-sized vector $\boldsymbol{h}_i \in \mathbb{R}^d$. The conditional intensity function $\lambda^*(t)$ is defined as a function of $\boldsymbol{h}_i$. We will use the transformer architecture (see previous lecture). We propose the following implementation.

Given the full sequence $\{t_1, t_2, \ldots, t_n\}$, we calculate all $\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n\}$ in parallel. We first calculate vectors $\boldsymbol{q}_i, \boldsymbol{k}_i, \boldsymbol{v}_i \in \mathbb{R}^d$ as a function of t_i. We stack these vectors into matrices $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{n \times d}$. The output of the transformer is: $\boldsymbol{H} = \text{softmax}(\boldsymbol{Q}\boldsymbol{K}^T)\boldsymbol{V}$, then $\boldsymbol{h}_i$ is the ith row of $\boldsymbol{H}$.

Identify the errors in this implementation compared to the original definition of $\boldsymbol{h}_i$. Propose a solution.

Problem 1: Consider a temporal point process, where all the inter-event times $\tau_i = t_i - t_{i-1}$ are sampled i.i.d. from the distribution with the survival function

$$S(\tau) = \exp\left(-\underline{(e^{b\tau} - 1)}\right) \quad = 1 - e^{b\tau}$$

with a parameter $b > 0$.

a) Write down the closed-form expression for the conditional intensity function $\lambda^*(t)$ of this TPP. Simplify as far as you can.

b) Write down the closed-form expression for the log-likelihood of a sequence $\{t_1, ..., t_N\}$ generated from this TPP on the interval $[0, T]$. Simplify as far as you can.

(a) $F^*(\tau) = 1 - S^*(\tau)$

$= 1 - \exp(-(e^{b\tau} - 1))$

$p^*(\tau) = \frac{dF^*(\tau)}{dt} = \cancel{-}\exp(-(e^{b\tau}-1)) \cdot (\cancel{-}e^{b\tau}) \cdot b$

$= \exp(1 - e^{b\tau} + b\tau) \cdot b$

$\lambda^*(t) = \frac{p^*(\tau)}{S^*(\tau)} = \frac{b \cdot \exp(1 - e^{b\tau} + b\tau)}{\exp(1 - e^{b\tau})} = b\,\frac{\exp(\cancel{1 - e^{b\tau}}) \cdot \exp(b\tau)}{\cancel{\exp(1 - e^{b\tau})}}$

$= b \exp(b\tau)$

(b) $p([0, T]) = \lambda^*(t_1) \cdots \lambda^*(t_N) \exp\int -\int_0^T \lambda^*(u)\,du$

$= b^N \exp\left(\sum_{i=1}^N bt\right) \exp$

intensity $h(\tau) = \lambda(\tau) = \frac{p(\tau)}{S(\tau)}$

$F(\tau) = 1 - S(\tau) \qquad p(\tau) = \frac{dF(\tau)}{d\tau}$

$= -\frac{\partial}{\partial\tau} S(\tau)$

$= -\frac{\partial}{\partial\tau} \exp(-(e^{b\tau} - 1))$

$= b \cdot e^{b\tau} \cdot e(-(e^{b\tau} - 1))$

$\lambda(\tau) = \frac{p(\tau)}{S(\tau)} = \frac{b \cdot e^{b\tau} \cdot \exp(-(e^{b\tau} - 1))}{\exp(-(e^{b\tau} - 1))}$

$= b \cdot e^{b\tau}$

$p(\{t_1, \dots, t_N\}) = \left(\prod_{i=1}^N f^*(t_i)\right) S(T)$

$= \left(\prod_{i=1}^N \lambda^*(t_i) S^*(t_i)\right) S^*(T)$

$\log p = \left[\sum_{i=1}^N \log \lambda^*(t_i) + \log S^*(t_i)\right] + \log S^*(T)$

$\log \lambda^*(t_i) = \log(b) + b(t_i - t_{i-1})$

$\log S^*(T) = 1 - \exp(b(t_i - t_{i-1}))$

$p = N \log b + N b (t_i - t_{i-1})$

$+ 1 - \exp(b(t_i - t_{i-1}))$

$t_0 = 0 \qquad t_{N+1} = T$

Problem 2: Consider an inhomogeneous Poisson process (IPP) on $[0, 1]$ with the intensity function $\lambda(t) = 2t$. We simulate a sample from this IPP using thinning. For this, we first simulate a *homogeneous* Poisson process (HPP) with intensity $\mu = 4$ and apply the thinning procedure described in the lecture. What is the expected number of events from the HPP that will be rejected when using this procedure?

$\mu = 4 \geq \lambda(t) = 2t$ in $[0, 1]$

assume simulate canadiacte events $\{t_1, \cdots, t_N\}$

keep probability $P_i = \frac{\lambda(t_i)}{\mu} = \frac{2t_i}{4} = \frac{1}{2} t_i$

Non - keep $\Rightarrow 1 - P_i$

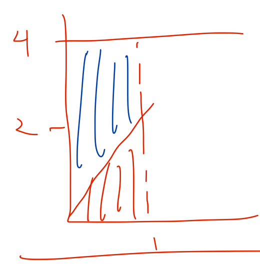
$E[R] = E(1 - P_i) \ldots$

$= 1 - E(P_i)$

$E(P_i) = \frac{1}{2} E(t_i)$

$\frac{1}{2} p \mu^2$

2 Hpp $\mu = 4$ Ipp $2t$



$\int_0^1 2t\,dt = \Big|_0^1 t^2$

reject 3 from 4

Problem 3: Consider an inhomogeneous Poisson process on $[0, 4]$ with the intensity function $\lambda(t) = \beta t$, where $\beta > 0$ is a parameter that has to be estimated. You have observed a single sequence $\{1, 2.1, 3.3, 3.8\}$ generated from this IPP. What is the maximum likelihood estimate of the parameter β?

$\lambda(t) = \beta t \quad [0, 4] \quad \beta > 0$

$p = \lambda(t_1)\lambda(t_2)\lambda(t_3)\lambda(t_4) \exp\left(-\int_0^4 \beta\mu\,d\mu\right)$

$= \beta^4 t_1 t_2 t_3 t_4 \cdot \exp(-[8\beta])$

$\frac{dp}{d\beta} = 4\beta^3 \_\_ \cdot \exp(-8\beta) \cdot (-8) = 0$

3 $\max_\beta \log p$

$\log\left(\prod_{i=1}^N \lambda(t_i)\right) \exp\left[-\int_0^T \lambda(\mu)\,d\mu\right]$

$= \sum_{i=1}^N \log \lambda(t_i) - \int^T \lambda(\mu)\,d\mu$

$= \sum \log \beta t_i - \int^T \beta\mu\,d\mu$

$= N \log \beta - \frac{T^2}{2}\beta + \underbrace{\sum \log t_i}_{\text{constant}}$

$\frac{\partial \cdot\cdot}{\partial\beta} = N\frac{1}{\beta} - \frac{T^2}{2} \overset{!}{=} 0$

$\beta = \frac{2N}{T^2} = \frac{8}{16} = \frac{1}{2}$

Problem 4: Consider a *neural* temporal point process where the conditional intensity function is defined with a neural network. In particular, for a time point t_i, we represent the history $\{t_1, t_2, \ldots, t_{i-1}\}$ with a fixed-sized vector $\boldsymbol{h}_i \in \mathbb{R}^d$. The conditional intensity function $\lambda^*(t)$ is defined as a function of $\boldsymbol{h}_i$. We will use the transformer architecture (see previous lecture). We propose the following implementation.

Given the full sequence $\{t_1, t_2, \ldots, t_n\}$, we calculate all $\{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n\}$ in parallel. We first calculate vectors $\boldsymbol{q}_i, \boldsymbol{k}_i, \boldsymbol{v}_i \in \mathbb{R}^d$ as a function of t_i. We stack these vectors into matrices $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{n \times d}$. The output of the transformer is: $\boldsymbol{H} = \text{softmax}(\boldsymbol{Q}\boldsymbol{K}^T)\boldsymbol{V}$, then $\boldsymbol{h}_i$ is the ith row of $\boldsymbol{H}$.

Identify the errors in this implementation compared to the original definition of $\boldsymbol{h}_i$. Propose a solution.

attention ?

RNN output the current state

Trans open u some previous knowledge

include all informations before hi

but RNN's hi-1 only contain hi-1 infor

$\lambda(t|H)$ vs $\lambda(t)$