Machine Learning and Data Mining - Homework 1

Philip Warton

October 2, 2021

1 Something

1.1 Q1

> Let $D = \{x_1, \cdots, x_N\}$ be a dataset from N poisson random variables, with a rate of $\lambda \in \mathbb{R}$. Derive the Maximum Likelihood Estimation for λ.

We begin by taking $\mathcal{L}(D)$. That is,

$$\begin{aligned}
\mathcal{L}(D) &= P(D \mid \lambda) \\
&= P(\{x_1, \cdots, x_N\} \mid \lambda) \\
&= P(x_1 \mid \lambda) \cdots P(x_N \mid \lambda) \\
&= \prod_{i=1}^{N} P(x_i \mid \lambda) \\
&= \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}
\end{aligned}$$

Then to get the log-likelihood, we take $\ln \mathcal{L}(D)$.

$$\begin{aligned}
\ln \mathcal{L}(D) &= \ln prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \\
&= \sum_{i=1}^{N} \ln\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right) \\
&= \sum_{i=1}^{N} \left[\ln(\lambda^{x_i} e^{-\lambda}) - \ln(x_i!)\right] \\
&= \sum_{i=1}^{N} \left[\ln(\lambda^{x_i}) + \ln(e^{\lambda}) - \sum_{j=1}^{i} \ln(x_j)\right] \\
&= \sum_{i=1}^{N} \left[x_i \ln \lambda - \lambda \ln e - \sum_{j=1}^{i} \ln x_j\right] \\
&= \ln \lambda \sum_{i=1}^{N} x_i - N\lambda - \sum_{i=1}^{N} \sum_{j=1}^{i} \ln x_j
\end{aligned}$$

We now take the derivative of the log-likelihood, giving us

$$\begin{aligned}
\frac{d}{d\lambda}\left(\ln \mathcal{L}(D)\right) &= \frac{d}{d\lambda}\left(\ln \lambda \sum_{i=1}^{N} x_i - N\lambda - \sum_{i=1}^{N} \sum_{j=1}^{i} \ln x_j\right) \\
&= \frac{1}{\lambda} \sum_{i=1}^{N} x_i - N - 0
\end{aligned}$$

Let this derivative be equal to zero. Then we have

$$
\begin{aligned}
0 &= \frac{1}{\lambda}\sum_{i=1}^{N} x_i - N \\
N &= \frac{1}{\lambda}\sum_{i=1}^{N} x_i \\
\lambda &= \frac{1}{N}\sum_{i=1}^{N} x_i \\
\lambda &= \overline{x}
\end{aligned}
$$

2 Something else

2.1 Q4

To see how these different embeddings of categorical variables in finite-dimensional Euclidean space, take, for example,