Churn Prediction Model Based on Beta Distribution

Description

This model is designed to predict user churn by modeling churn time (T) as a discrete random variable, influenced by a hidden continuous rebilling probability (p), which follows a Beta distribution.

General Loss Function

In the problems like survival modeling we always have the share of data for which we do not know the outcome T, but we know that $T > t$. In such a case, we should incorporate this information through censoring. We use $c \in \{0, 1\}$ for noting censoring ($c = 1$ means that T is not observable), t_i for the last observable period of a user i. Then the likelihood will be the following:

$$\prod_{i=1,c_i=0}^{n} P(T_i = t_i) \prod_{i=1,c_i=1}^{n} P(T_i >= t_i)$$

In discrete case we have:

$$\prod_{i=1,c_i=0}^{n} P(T_i = t_i) \prod_{i=1,c_i=1}^{n} S_i(t_i - 1)$$

where $S_i(t_i - 1) = P(T_i > t_i - 1)$.

Taking the log we obtain

$$\mathcal{L} = \sum_{i=1}^{n} log(P(T_i = t_i))\mathbb{I}_{c_i=0} + \sum_{i=1,}^{n} log(S_i(t_i - 1))\mathbb{I}_{c_i=1}$$

where $\mathbb{I}_x$ is an indicator function.

Base Model

Rebilling Probability

$p \in (0, 1)$ — a continuous hidden random variable representing the probability of a user rebilling.

Churn Time

$T \in \{0, 1, 2, 3, 4, 5, 6, ...\}$ — the discrete time period at which churn occurs.

Probability of Churn at Time t for Fixed p

$$P(T = t \mid p) = p^t \cdot (1 - p)$$
$$P(T > t \mid p) = p^{t+1}$$

Prior Distribution for p

$$p_i \sim \text{Beta}(\alpha(x_i), \beta(x_i))$$

Marginalized Distribution

$$\begin{aligned}
P(T_i = t_i \mid \alpha(x_i), \beta(x_i)) &= \int_0^1 P(T_i = t_i \mid p)\, f(p \mid \alpha(x_i), \beta(x_i))\, dp \\
&= \int_0^1 p^{t_i}(1-p) \cdot p^{\alpha(x_i)-1}(1-p)^{\beta(x_i)-1}\, dp \\
&= \int_0^1 p^{t_i+\alpha(x_i)-1}(1-p)^{\beta(x_i)}\, dp \\
&= \frac{B(\alpha(x_i) + t_i, \beta(x_i) + 1)}{B(\alpha(x_i), \beta(x_i))}
\end{aligned}$$

Censored Observations

$$P(T > t_i \mid \alpha(x_i), \beta(x_i)) = \frac{B(\alpha(x_i) + t_i + 1, \beta(x_i))}{B(\alpha(x_i), \beta(x_i))}$$

Then

$$S_i(t_i - 1 \mid \alpha(x_i), \beta(x_i)) = \frac{B(\alpha(x_i) + t_i, \beta(x_i))}{B(\alpha(x_i), \beta(x_i))}$$

Logarithm of the Beta Function

$$\log B(\alpha, \beta) = \log \Gamma(\alpha) + \log \Gamma(\beta) - \log \Gamma(\alpha + \beta)$$

Loss Function

The final loss function for the model, which predicts $(\alpha(x_i), \beta(x_i))$, considers both observed and censored data:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [\log B(\alpha(x_i) + t_i, \beta(x_i) + 1 - c_i) - \log B(\alpha(x_i), \beta(x_i))]$$

where:

- t_i is the observed churn time
- c_i is the censoring indicator (1 if the observation is censored, 0 otherwise)

Beta-Discrete-Weibull Model

The Beta-Discrete-Weibull (BdW) model extends the basic Beta model by introducing a Weibull scaling factor $\gamma(x_i)$.

Conditional Distribution

$$P(T = t \mid p) = p^{t^\gamma} - p^{(t+1)^\gamma}$$

$$P(T > t \mid p) = 1 - \sum_{\tau=0}^{t} P(T = \tau | p) = p^{(t+1)^\gamma}$$

Prior Distribution

$$p_i \sim \text{Beta}(\alpha(x_i), \beta(x_i))$$

Marginalized Distribution

$$P(T_i = t_i \mid \alpha(x_i), \beta(x_i), \gamma(x_i)) = \frac{B(\alpha(x_i) + t_i^{\gamma(x_i)}, \beta(x_i)) - B(\alpha(x_i) + (t_i + 1)^{\gamma(x_i)}, \beta(x_i))}{B(\alpha(x_i), \beta(x_i))}$$

Censored Probability

$$P(T_i > t_i \mid \alpha(x_i), \beta(x_i), \gamma(x_i)) = \frac{B(\alpha(x_i) + (t_i + 1)^{\gamma(x_i)}, \beta(x_i))}{B(\alpha(x_i), \beta(x_i))}$$

$$S_i(t_i - 1 \mid \alpha(x_i), \beta(x_i), \gamma(x_i)) = \frac{B(\alpha(x_i) + t_i^{\gamma(x_i)}, \beta(x_i))}{B(\alpha(x_i), \beta(x_i))}$$

Log-Likelihood Function

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n} \Big[\log \Big(B(\alpha(x_i) + t_i^{\gamma(x_i)}, \beta(x_i)) - B(\alpha(x_i) + (t_i + 1)^{\gamma(x_i)}, \beta(x_i)) \Big) \cdot \mathbb{I}_{c_i=0} \\ + \log B(\alpha(x_i) + t_i^{\gamma(x_i)}, \beta(x_i)) \cdot \mathbb{I}_{c_i=1} - \log B(\alpha(x_i), \beta(x_i)) \Big]\end{aligned}$$