On Time Series

Randolf Scholz

June 23, 2021

Definition 1 (time series). A *time series* S over a *time space* $\mathcal{T}$ and a *data space* $\mathcal{X}$ is a finite sequence $S = (t_i, x_i)_{i=1:|S|} \in (\mathcal{T} \times \mathcal{X})^*$ such that

1. The time space $\mathcal{T}$ is an additive, totally ordered *monoid* satisfying $\forall a, b \in \mathcal{T} : a \leq b \implies \exists c \in \mathcal{T} : a + b = c$

2. The time stamps are in order $t_i \leq t_j \forall i \leq j$

3. The data space $\mathcal{X}$ is a *set* containing a distinguished element `NaN`, which we will use to denote missing (unobserved) values

Definition 2. We say a time series is

1. *regular*, iff $\exists \Delta t > 0 \forall i : \Delta t_i = \Delta t$

2. κ-*quasi-regular*, iff $\exists \Delta t > 0 \forall i \exists k \in \mathbb{N}_0 : \Delta t_i = k \Delta t$. In this case, given $\Delta t_{\min} = \min\limits_{\Delta t_i > 0} \Delta t_i$, and $\Delta t_{max} = \max\{\Delta t > 0 : \forall i \exists k \in \mathbb{N}_0 : \Delta t_i = k \Delta t\}$ we call $\kappa = \frac{\Delta t_{\min}}{\Delta t_{\max}}$ the *regularity quotient*. κ behaves similar to a condition number: $\kappa \in [1, \infty)$, the larger κ the farther from regular the series is. One will have to insert up to $\kappa |S|$ dummies to make the TS regular. Note that for integer data, $\Delta t_{max} \gcd(\{\Delta t_i > 0\})$

3. *irregular*, iff it is neither regular nor quasi-regular.

Generally, an irregular time series can be made regular by inserting *enough* "empty" observations consisting only of `NaN`'s. Note that this is true only practically, since there is an automatic discretization through the use of floating point numbers.

Therefore, the definition of quasi-regular does not cover simple cases like $(2, 4, 6, 8, 11, 14, 17, 20, 22, 24, 26)$ with timedeltas $2, 2, 2, 2, 3, 3, 3, 3, 2, 2, 2$. Here, we could insert a single dummy whenever the timedelta is 2 and two dummies whenever it is 3 to get a regular time-series. However, when the time-steps are like $2, 4, 6, 8, 8 + 2^{-10}$ then it makes little sense to fill in the gaps since it would require us to fill in so many dummies that we would drown the signal.

In practice, since floating-point numbers are essentially rational numbers in the localization $\mathbb{Z}(-2)$, every time series will be quasi-regular. Thus in practice, we ought to decide at which point we do a cut-off

Definition 3 (ε-gcd)**.** We define the *approximate greatest common divisor* (ε-gcd) of a set of n numbers $x_1, \ldots, x_n$ as

$$\varepsilon - \gcd(x_1, \ldots, x_n) = \max\{y \mid \forall i \exists k \in \mathbb{Z} : x_i \in U_\varepsilon(k \cdot y)\} \tag{1}$$

$$= \max\{y \mid \forall i \exists \operatorname{dist}(x_i, y\mathbb{Z}) \leq \varepsilon\} \tag{2}$$

Example 4 (title)**.** COnsider the time steps $(1, 2, 3, 4, 4 + 2^{-10})$. This has a strict floating-point gcd of 2^{-10}. However, in a situation like this it may be most appropriate to not treat it as an irregular time series, not an quasi-regular time-series, but rather just merge the last two datapoints (e.g. by averaging). To deal with these cases, we generalize the notion of the floating-point gcd to a general real gcd, which is based on a notion of approximate equality.
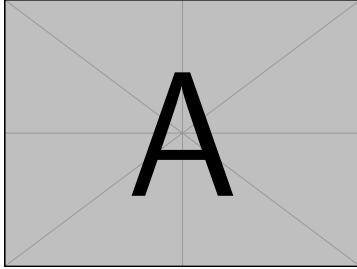
1 Encoding of Features

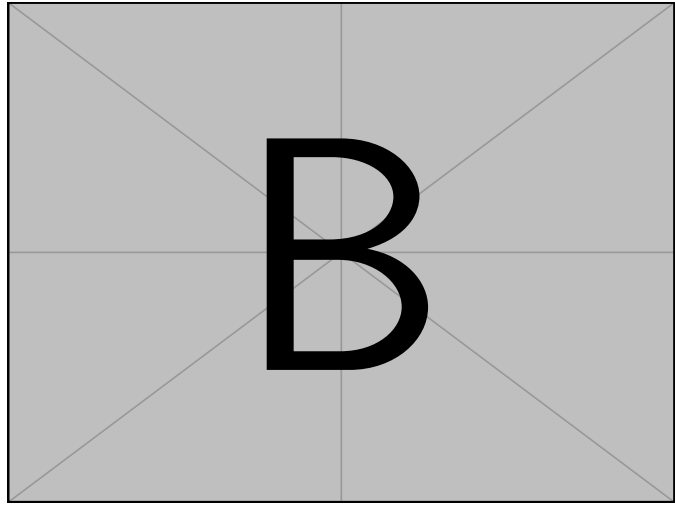Categorical Features: One Hot

2 Filter Component

Figure 1: left

Figure 2: right