2/28 → 2/27 ddl   2/11 – 2/27   14天.

1. 收集数据   2/11 – 2/13   13晚   美东晚八.
2. EDA   2/13 – 2/15.   15晚
3. 正式数据   2/15 – 2/17
4. 改善   2/17 – 2/19
5. 确认   2/19 – 2/21
6. Narrative   2/21 – 2/23
7. Code submit   2/23 – 2/25.

tempre →

Python

+1 • Template
+2 • 简历

-1 • 统计包 ×

R ✓

+1 • TS / forecast / 包 ✓.
+2 • Rmd.

1st

Prof. Japanese Language Culture

Cultural significance of cherry & cherry blossom in Japan.

flower viewing - peak blooming time

late march ~ early April

Map 2017~2021

2nd.

phenological model

Timing

· Harvest date

global warmth → shift a month early

·

{ 1731 Reamur scale　0 — 100 water
　freeze. boil
1735 GDD

① Global warm vs. local temperature

② GDD +winter chilling model

③ Empirical Function

$TSD = \exp\{ \cdots \frac{T-0}{288.5T} \}$

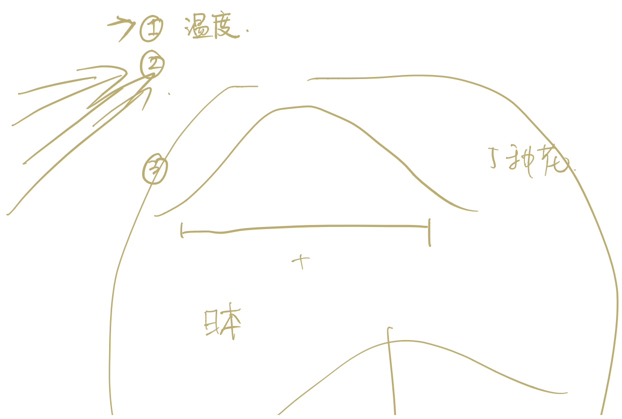
④ 花开速度

bucket model

< winter chilling
　spring warming GDD

crop models (stone fruits)
wild animals
forest ecosystems

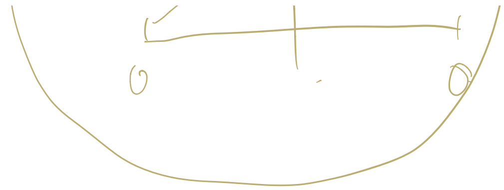
① 温度　昼夜
② 太阳　↑↓
③ 云覆盖率
④ 湿度
⑤ 海拔



+ 下载 + 学习 Rstudio.

3rd

nature's notebook



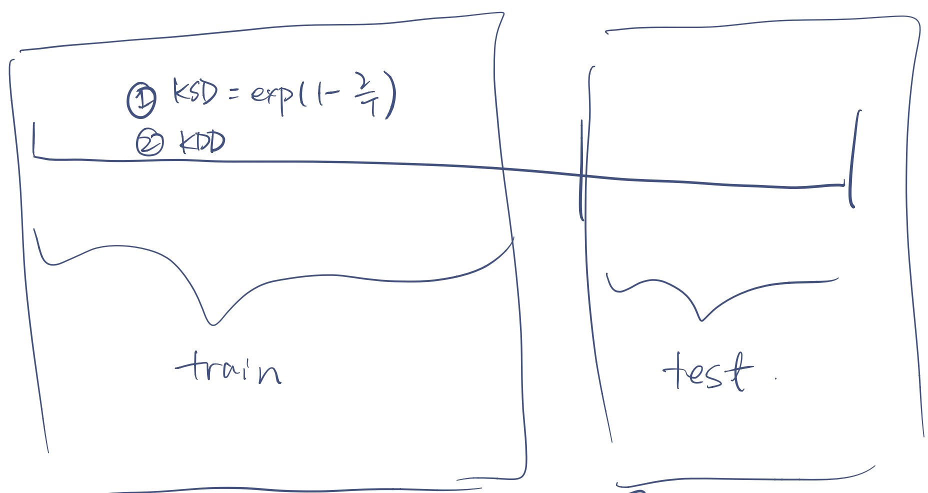
USA-NPN

五种花 2009-2021 的数据

status & intensify

individual ID → plants ID

individual phenometrics

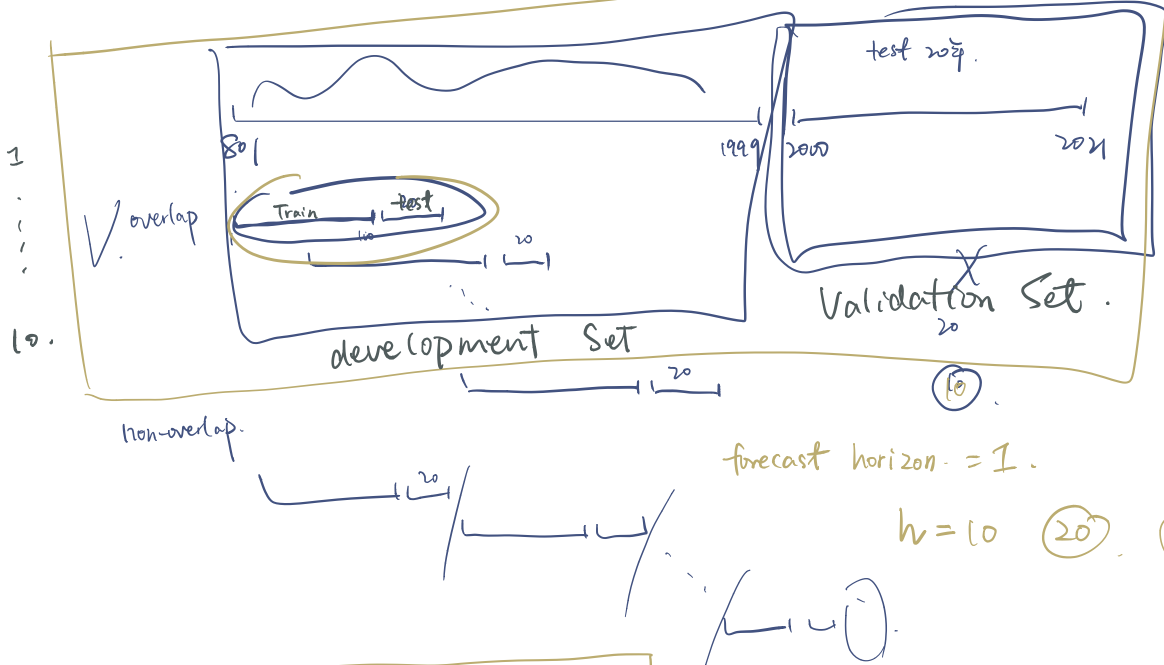


① T → KSD → 1天

② → 2天

10-fold cross validation

Kyota 801 - 2021 804. bloom.



forecast horizon = 1.

h = 10 20 1

KDD model. (过去20年).

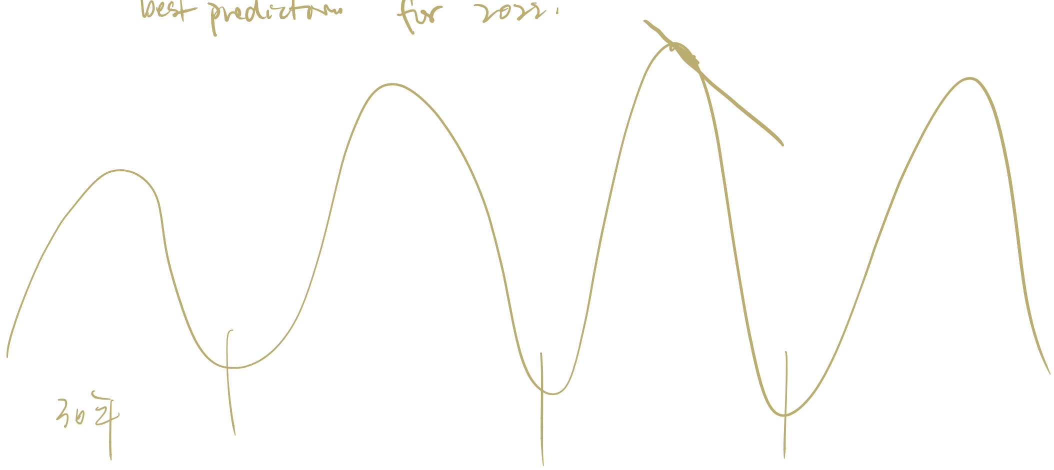
20年. 10年

2022 — 2031

KDD(2001 - 2021) → 2022-2031.

KDD(1980 - 2000) → 2001-2011

best prediction for 2022.



$$KSD = \exp\left\{ 1 - \frac{T-288.5}{288.5T} \right\} \rightarrow \text{Japan} . \quad X .$$

$$\log KSD = 1 - \frac{1}{288.5} - \frac{1}{T} \approx 1 - \frac{1}{T}$$

KDD. $T + S$.

2022/02/13

1、	收集数据	2/11 - 2/13	13晚
2.	EDA	2/13 - 2/15 .	14 - 15
3.	正式数据	2/15 - 2/17	16-17
4.	改善	2/17 - 2/19	18-19
5.	确认	2/19 - 2/21	
6.	Narrative	2/21 - 2/23	
7.	Code submit	2/23 - 2/25.	

瑞士　小时

日本　附近
　　　月为单位
　　　pdf 非 csv.

加拿大　气温

DC　station

日本 ①	DC
瑞士 .	加拿大

① 日本　T= Temperature

winter chill　summer warm

T → GDD_WC ← V1
　　GDD_SW ← V2.

pt — GDD → GDD f1

compare 经验公式 with f1.

$y = \beta_0 + \beta_1 V_1 + \beta_2 V_2$ 日本.

y = logit

exp

f1 → 日本

日本. 日照.

DC. 云浩

. . . . . .

Y

② .

β_0 β_1 β_2

V_3 →

V_4.

模型化: ① $l_1 = lm(y \sim x)$ data = Japan

⇒ residual plots (l_1)

data cleaning
1. feature
2. missing value (mean → Amelia / Mice)
3. weight . . .

总结.

日本. . . $\exp(\hat{\beta}_0 + \hat{\beta}_1 V_1 + \hat{\beta}_2 V_2)$

Hochmaier 2014

$T_0 = 4.5°C$

$GDD = \sum_{t=t_0}^{t_1} (T - T_0)$

CDD $\boxed{\text{Dormancy time}}$

1050 ~ 1900 h < 7°C 1982~1985

1081 ~ 1214 h 3.2-3.7°C 2000

• 1200 h $\leq$ 6°C at → 2008

$\boxed{\text{Base temperature } T_0 = 4.5°C}$

4°C Sour cherry 1980

• 4.5°C Iezzoni 1985 ← bloom

$\boxed{\text{Chilling Unit}}$

Temp (°C)	<2.5	2.5~9.2	9.2~12.5	12.5~16	16~18	>18
CU	1	1	0.5	0	−0.5	−1

Empirical CU 1615 (SD = 161)
1550 (SD = 118) ⇒ CU ≈ 1600 ± 150

$\boxed{\text{Degree-Day}}$

$CDD = \sum_{t=101}^{t} (T - T_0)$ ⇒ 多少.

Bing	Van	Lapins	Sweetheart
805 (75)	765 (66)	852 (64)	916 (88)

1. Avg t − T_0

2. $\frac{\text{low t} + \text{high t}}{2} - T_0$

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

t_0 starting day of CDD $\boxed{t_0 = 1\,?}$

T_0 $4.5°C$

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom satisfy the requirement.

When $CDD_t \geq CDD_{th} \Rightarrow t_b$ blossom time

$$t_b = \underset{t}{\operatorname{argmin}} (CDD_t > CDD_{th})$$

Time series model for temperature estimation T_t

$\Leftrightarrow$ - - - - - - - - - - - - - - CDD_t

Assume T_t follows $AR(p)$

$$p_t = P(CDD_t > CDD_{th})$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 T_t$$

数据.

$z_t = 1$ If $CDD_t \geq CDD_{th}$

$z_t = 0$ If $CDD_t < CDD_{th}$

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

t_0 starting day of CDD $\boxed{t_0 = 1\,?}$

T_0 $4.5°C$

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom satisfy the requirement.

When $CDD_t \geqslant CDD_{th} \Rightarrow t_b$ blossom time

$$t_b = \underset{t}{\operatorname{argmin}} \left(CDD_t > CDD_{th} \right)$$

Time series model for temperature estimation T_t

$\Leftrightarrow$ - - - - - - - - - - - - - - CDD_t

Assume T_t follows $AR(p)$

$$p_t = P(CDD_t > CDD_{th})$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 T_t$$

数据.

$z_t = 1$ If $CDD_t \geqslant CDD_{th}$

$z_t = 0$ If $CDD_t < CDD_{th}$

Simple LR $CDD_t = \beta_0 + \beta_1 T_t$

location = ℓ = koyota

Cumulative degree day

$$CDD_t = \sum_{t=t_0}^{t} (T_t - T_0)$$

$t_0 = 1$ starting day of CDD $\boxed{t_0 = 1\,?}$ ①

T_0 4.5°C

T_t Average daily temperature of day t

CDD_{th} Threshold CDD_{th} at which cherry blossom

800 satisfy the requirement. ②

When $CDD_t \geqslant CDD_{th} \Rightarrow t_b$ blossom time

T_t. $t_b = \underset{t}{\operatorname{argmin}} (CDD_t > CDD_{th})$ (1)

Time series model for temperature estimation $\boxed{T_t}$

$\Leftrightarrow$ - - - - - - - - - - - - - CDD_t

Assume T_t follows AR(p) ③ ZXY.

$\boxed{P_t = P(CDD_t > CDD_{th})}$

$P_t \in (0,1)$ $\frac{(0,1)}{(0,1)}$

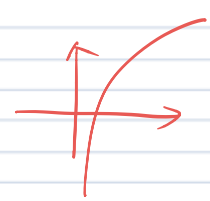
$\frac{P_t}{1-P_t} \in (0, +\infty)$

$$\text{logit}(P_t) = \log\left(\frac{P_t}{1-P_t}\right) = \underbrace{\beta_0 + \beta_1 T_t}_{(-\infty,+\infty)} \quad (2)$$

[illegible].

$Z_t = 1$ If $CDD_t \geqslant CDD_{th}$

$Z_t = 0$ If $CDD_t < CDD_{th}$



$\log\left(\frac{P_t}{1-P_t}\right) \in (-\infty, +\infty)$

$P_t \in (0,1)$

intercept slope

$$\text{logit}(P_t) = \beta_0 + \beta_1 T_t$$

train.

<ii> $T_t \sim AR(p)$

<i> $Z_t. = T_t \rightarrow \hat{\beta}_0 \ \hat{\beta}_1$

K1Train $Z_t, T_t. \rightarrow \hat{\beta}_0 \ \hat{\beta}_1$

K1Test $\cancel{Z_t}$. $\hat{Z}_t = \hat{\beta}_0 + \hat{\beta}_1 T_t$

$Z_t \ \hat{Z}_t \rightarrow$ Accuracy.

M_B. Best model $= \underset{\beta.}{\arg\min} \ Err(Z_t, \hat{Z}_t)$

<iii> M_B.

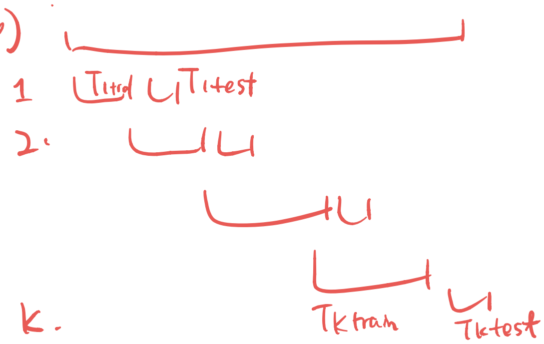
T_{test}

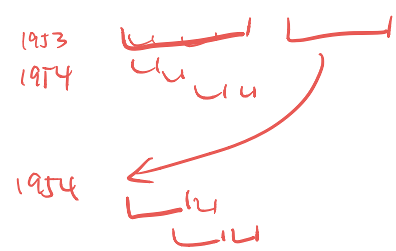
$\hat{Z}_{test} = M_B \cdot T_{test} \rightarrow$

glm($Z_t \sim T_t$, link = "binomial")

(a) CDD_t

(b)





(c) $Z_t = CDD_t \geq CDD_{th}$

$Z_t, T_t \rightarrow$ model (2) logit.

k上 train $\hat{\beta}_k$ $\rightarrow$ k上 test $\hat{z}t_k$ 

$\hat{p}_t$

2022/02/15

$$CDD_t = \sum_{t=t_0}^{t} (T - T_0)$$

t_0 起始时间

T_0 4.5°C temperature

$$\Delta_t = CDD_t - CDD_{th}$$

① $= \sum_{t=t_0}^{t} (T - T_0) - CDD_{th}$

② $= \sum_{t=t_0}^{t} T - \left[(\underset{\Delta}{t} - t_0) T_0 + CDD_{th} \right]$

$= \sum_{t=t_0}^{t} T - T_0 \cdot t + t_0 T_0 + CDD_{th}$

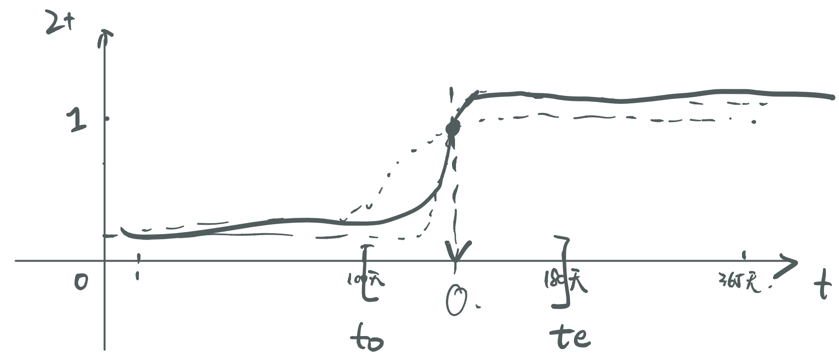
cumsum(Avgtem) time

$p_t = P(CDD_t > CDD_{th})$ ✗

京都

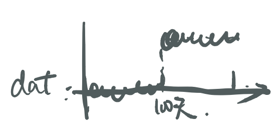
$$\begin{cases} Y_t = 1 & \geq 80\% \text{ blossom } ✓ \\ Y_t = 0 & < 80\% \text{ blossom } ✗ \end{cases}$$

day1 day365



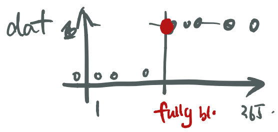
$CDD_t = \sum_{t=t_0}^{t} T_t$, $P_t = $ Prob(fully blossom at time t) "response"

$$Y_t \quad \text{logit}(p_t) = \log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 CDD_t + \beta_2 \cdot t + \varepsilon.$$ Model. (1)



$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$Y = X\beta$



$\hat{\beta} = (X'X) X' Y$

$\hat{Y} = X\hat{\beta}$

$$\log\left(\frac{p_t}{1-p_t}\right) = A = \beta_0 + \beta_1 CDD_t + \beta_2 t$$

$$\frac{p_t}{1-p_t} = \exp\{A\} = e^{A}$$

$$\frac{1-p_t}{p_t} = e^{-A}$$

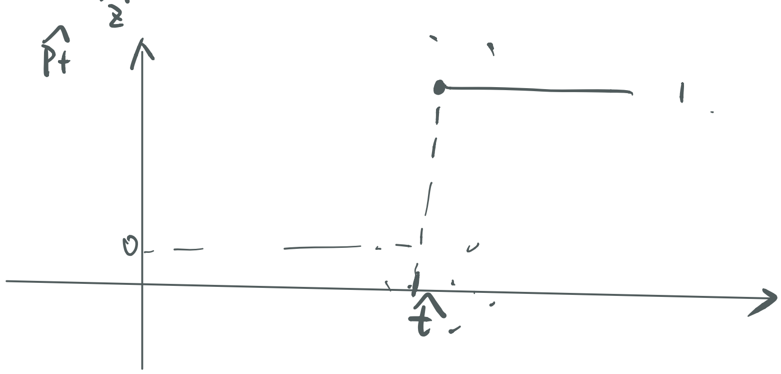
$$\frac{1}{p_t} - 1 = e^{-A}$$

$$\frac{1}{p_t} = 1 + e^{-A}$$

$$p_t = (1+e^{-A})^{-1}$$

$\hat{z}_t$

$$p_t = \left(1 + \exp\{-\beta_0 - \beta_1 CDD_t - \beta_2 t\}\right)^{-1} \quad \text{Model (2)}$$



glm(y= , x= $CDD_t + t$).

2/17 晚.

<i> 四国家 impute data

<ii> 给出预测 Development Accuracy

1个train test $\left\langle \begin{matrix} t_b \\ \hat{t}_b \end{matrix} \right.$

Define forecast error

Model Comparison

RMSE $= \sqrt{\sum_{i=1}^{3} (t_b - \hat{t}_b)^2}$

<iii> TS 模型

	logit(p_t) ① $\beta_1 CDD_t + \beta_2 t$	② $\beta_1 \log CDD_t$	经验公式.
RMSE.	100	50.	

```
glm( model= arg.----.            ).

f(  model.                 )

f( ) <- f( model. arg              ) {

  没
  分

   glm

   accuracy

} return (. list ( model= ..., accuracy ) )
```

TS Split

$k=100$.

Train : $K = 100$
Test : h

$t_1 \quad t_2 \quad - - - - - - \quad t_n$

Overlap

① $\underbrace{1 \quad 2 \quad \cdots \quad k}$ $\underbrace{k+1 \quad \cdots \quad k+h}$ $h=1$ $k+1$天 ← l_1.

② $\underbrace{2 \quad \cdots \quad k+1}$ $\underbrace{k+2 \quad \cdots \quad k+h+1}$ $k+2$天.

$n-h-k+1$

80.

$\underbrace{n-h-k+1 \quad \cdots \quad n-h}$ $\underbrace{n-h+1 \quad \cdots \quad n}$ n天.

Approch 1

$l_1($ 101天 $\begin{matrix} z_{101} \\ Y_{101} \end{matrix}$ $\sim$ [1~100天 t / T] $) \rightarrow \hat{z}_{101}$

Approach 2:

$$Y = \begin{bmatrix} p_{k+1} \\ p_{k+2} \\ \vdots \\ p_n \end{bmatrix} \leftarrow \text{Train} \begin{bmatrix} t_1 & \cdots & t_k \\ t_2 & \cdots & t_{k+1} \\ \vdots & & \vdots \\ t_{n-h-k+1} & \cdots & t_{n-h} \end{bmatrix}$$

$\underbrace{(n-h-k+1)}_{\text{行}} * k$ 列

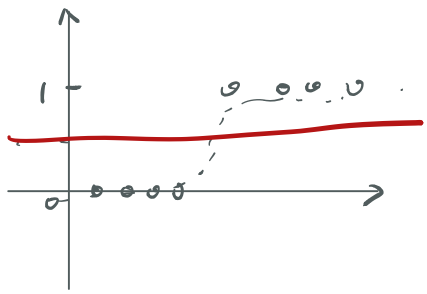
$X_{train} =$

$Y=[\ ] \sim X_{train}$

$n-h-k+1$ estimate ← pred( ).

$Y - \hat{Y} =$ $n-h-k+1$ 个 预测 error.

分类模型 logistic



k+1天. True.

False

实际 \ 预测	$\hat{z}$	$\hat{z}=1$
$z=0$ 未开放.		
$z=1$ 开放		

① · Time series $T \sim AR(\ )$.

② $t_b = \exp\left\{1 - \frac{305-T}{T}\right\}$. + location

Given Validation T. $\hat{t_b}$.

glm(CDD+ ↓ 温度. t. ↓ 时间).