Imb-Data+DT
Mon, 9 Aug 2021 10:30

Agenda :

- Imbalanced data
- Impact
- Handle Imb data
- DT Intro.
- Geometric Intuition
- Entropy
- Information Gain
- Gini
- Code - DT

Imbalance data

per class → distribution

$0_{(+)} : 1_{(-)}$ — Balanced

$0_{(+)} : 100_{(-)}$ — $\frac{1}{101}$ Very Imbalanced

$99_{(+)} : 1_{(-)}$ — $\frac{99}{100}$ Very Imb.

$100_{(+)} : 1_{(-)}$ — Imb data

$99_{(+)} : 1_{(-)}$ — extremely imb.

Q: what is the impact?

$\boxed{k=3}$

$\boxed{k=25}$

$\boxed{k=51}$

as $\boxed{k \uparrow}$ data imb. will impact more known.

if $\boxed{k \downarrow\downarrow}$ $k=1$

Imb-Data with logistic Reg.

$\log_{loss} = \sum_{i=1}^{n} \log loss_i$

Handle Imb Data :

① class-weights. (w) $w_i \; (i=1:n)$

$\sum_{i=1}^{n} \log loss_i * w_i$

Ex → loss 4.5 ✓ (+) loss (−)

$\frac{80}{20} = \boxed{4}$

majority minority → (w) → 4 for majority, 1 for minority

$4 \times 6 : y_i = 1 \; (+)$

$4 \times 1 : y_i = 0 \; (-)$

$\frac{1}{2}$

$\frac{1}{2}$

$\frac{1}{2}$