Pakdd-2021-ToT

A) DFN of anomaly
B) Automation?

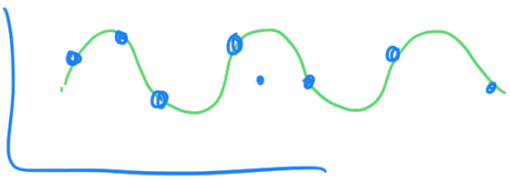
A)

outlier depends on our explicit/implicit biases/(statistical) model
↖ Gaussian/unif ...

B) "N."

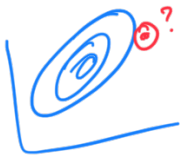
anom ↔ (B1) pattern → (B4') compression → Kolm compl UNDECIDABLE
anom → (B2) dist ↔ (B3) density → (B4) compression

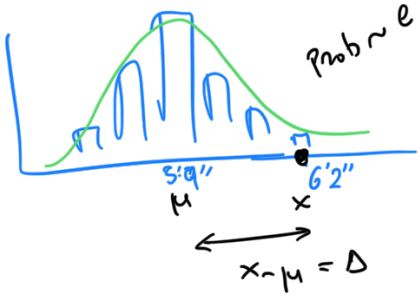
B1 anom ↔ pattern



B2 anom ↔ dist

B3) dist ⇝ density (≃ prob)





$-\log \text{prob} \propto \Delta^2$

B4) density ⇝ compression

density ~ prob

½ H
½ T

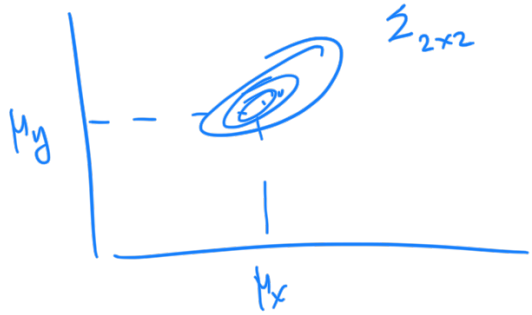
code ('H') → $-\log_2 1/2 = 1$

A → 1/8
B : "
H : 1/8

code ("A") → $-\log 1/8 = 3$

$-\log P_A \simeq$ coding cost

B4') patterns ⇝ compression

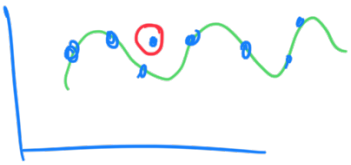


Kolm compl : undecidable

But, practically:

- make assumptions
  = (stat.) models (Gauss; Poisson; ...
  (physics) models $\ddot{x} = -kx$
  (other) "

- best model → best compression
  → (few) outliers



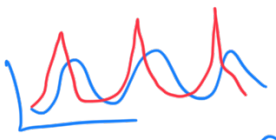
(FRACTALS?)

Q: How to find "good" models?

A: cross-disciplinarity

EEG FHN

pop ecol L.V.



chaos