Intro-data-DT

Monday, December 12, 2022 10:12 PM

Q → Class de sonido

"Es bueno"

Agenda :

→ Introducirnos data.
→ Entropy
→ Modelo Info Data
→ DT Intro.
→ Semantic Intuition
→ Entropy / IG } → mathematics

Introducirnos data.

one class → domination

$y_i = 0$ → Robotic

$y_i = 1$ → Slightly Robotic

$y_i = y_j$ → Semantic robotics

$y_i \neq y_j$ → Individual

$y_i = y_j$ → actually Sub.

Indicar → Nivel

$\boxed{k=1}$

$\boxed{k=2}$

⋮

$\boxed{k=k}$

as $\boxed{k=1}$ Info data will indicar more a more.

If $k \leq k$; $\boxed{k << n}$ (less indicar of info data)

Q: How to find k_{max}?

Indicar → logaritmo equivalente.

$$\log_2 x \Leftrightarrow \sum_{i=1}^{n} \log_2 \log_2 \; \text{entry}$$

Manejo Info. Data.

① Class weights.

$$\log(w) \Leftrightarrow \sum_{i=1}^{n} \log \log_2 \times w_i$$

$w_i = 1/\text{Volumen}$

Data → $y_i = 0$ (-ve) ; $y_i = 1$ (+ve)

$y_i = 0$ → $\sum_{i=1}^{n} y_i = 6$

$y_i = 1$ → $\sum_{i=1}^{n} y_i = 0$

$\frac{6}{6+0} = 6$

$w_i^{-ve} = 6$; $y_i = 1$ (+ve)

$w_i^{+ve} = 1$; $y_i = 0$ (-ve)

mayoría → minoría

minoría → mayoría

6:1

class_weight = { 0 : 1, 1 : 6 }

② Oversampling

Random oversampling

$\frac{1000}{10} = 100$ → +ve

+ 900 → +ve

$\frac{10}{10} = 1$ → -ve

$\sum_{i=1}^{n} \text{Ejemplos de clase minoría}$

$i = 0$

③ Undersampling

Randomly sample the majority class ve class

$\frac{1000}{100} = 10$ → +ve

100 → -ve

$\sum_{i=1}^{n} \text{majoría} - \text{Ejemplos}$

$i = 0$

Why are you deleting data?

Avoid large w. → $\frac{1}{\text{Volumen}}$

$\frac{1000}{100} = 10$ → +ve

100 → -ve

Undersampling.

$w_i^{+ve} = \frac{1000}{100}$

SMOTE [Synthetic Minority Oversampling Technique]

→ Data Augmentation.

→ Nuevo

$\boxed{k=5}$

① Randomly fix a point "x"

② for i in range(k):

$r \leftarrow U(0,1)$

$\text{nuevo} = x^{(1)} + r \cdot \text{dist}(x^{(1)}, x^{(i)})$

let $r = 0.2$

$\text{nuevo} = x^{(1)} + 0.2 \cdot \text{dist}(x^{(1)}, x^{(i)})$

$\text{Dist}(x_1, x_2)$

"Tomar una dist" de la distancia entre cada vecino (x_1, x_2)

$x_1 \cdot x_2$

$\frac{x_1 \cdot x_2}{\|x_1\| \cdot \|x_2\|}$

$\frac{x_1 \cdot x_2}{\|x_1\| \cdot \|x_2\|} = \frac{x_1}{\|x_1\|} \cdot \frac{x_2}{\|x_2\|}$

Decision Tree.

→ Decision Tree

→ Semantic Main

→ Entropy : RE or "Ambiguity"

→ Knowledge : GRAFO → "Algoritmo de Dijkstra" (Graph)

"Entropía → Aleatorio"

① Random Walk

② RE → "Noise"

DT Intuition

→ Determinar Algo. (+,-)

→ Semantic Mp (Conclusiones) Knowledge

→ Cross Validation / Bagging

$\uparrow$ → Tener

$\downarrow$ → Vivir

If $\boxed{\text{age} < 18}$:

If $\boxed{\text{another } < 5}$:

→ $\boxed{\text{Y}}$

else:

→ $\boxed{\text{Lame}}$

Else:

If $\boxed{\text{age} < 50}$

→ $\boxed{\text{Y}}$

else:

→ $\boxed{\text{Lame}}$

Decision Tree

Tu tarea : Condición / Nodo

option a (Clase)

option b

"Mover Feat Nodo"

"Mover Impureza"

Objetivo :→ Get Condiciones where you get more "Pure nodes"

→ Quantify | measure "Impureza" or ambiguity.

Entropy

→ Entropy = "Impureza"

→ Ambiguity

→ Info. Aleatoria

→ "Randomness"

$$Y : \text{Objetivo a } \leftarrow \quad Y \in \{y_1, y_2, y_3, \dots y_n\}$$

$$Y \in \{+, -\}$$

$$\boxed{H(Y) \rightarrow -\sum_{i=1}^{k} p(y_i) \cdot \log_2(p(y_i))}$$

entropy

$Y \in \{+, -\}$

$0, 1$

$$H(Y) \rightarrow -\left[p(+) \cdot \log p(+) + p(-) \cdot \log(p(-))\right]$$

$p(y=+) \rightarrow p$

$p(y=-) \rightarrow 1-p$

$$\boxed{H(Y) \rightarrow -p \log p - (1-p) \log(1-p)} \rightarrow \text{Base } 2 \text{ log-law}$$

Binary log law $\Rightarrow -y \cdot \log y - (1-y) \cdot \log(1-y)$

Binary classification

① Data : y_i : 1ev/.

y_i : 0ev/.

$$\Rightarrow -\left[p(1) \cdot \log(p(1)) + p(0) \cdot \log(p(0))\right]$$

$$\Rightarrow -\left[0.5 \times \log 0.5 + 0.5 \cdot \log 0.5\right]$$

$\Rightarrow \boxed{1}$

② Data : y_i : 1ev/.

y_i : 0ev/.

$$\Rightarrow -\left[p(1) \cdot \log(p(1)) + p(0) \cdot \log(p(0))\right]$$

$$\Rightarrow -\left[0.9 \cdot \log 0.9 + 0.1 \cdot \log 0.1\right]$$

$\Rightarrow \boxed{0.469}$

③ Data : y_i : 1ev/.

y_i : 0ev/.

$$\Rightarrow -\left[p(1) \cdot \log(1) + p(0) \cdot \log(0)\right]$$

$$\Rightarrow -\left[1 \cdot \log(1) + 0 \times \log(0)\right]$$

$0 \quad 0$

$\Rightarrow \boxed{0}$

Binary class

Entropy.