RICAPITOLO

POSSIAMO RIASSUMERE:

- VARIABILI:
  - $a_t \in \{0,1\}^p \longrightarrow$ coefficiente attributo test
  - $b_t \in [0,1]$, $\forall t \in T_B \longrightarrow$ termine noto sul test
  - $d_t \in \{0,1\} \longrightarrow$ c'è split?
  - $z_{it} \in \{0,1\} \longrightarrow$ il punto $x_i \in t$?
  - $|T_{BRANCH}| = 2^D - 1$, $|T_{LEAF}| = 2^D$
- VARIABILI AUSILIARIE:
  - $c_{kt} \in \{0,1\}$, $k=1 \dots K \longrightarrow$ la classe k è quella di t?
  - $l_t \in \{0,1\}$, $l_t \in \mathbb{Z}$, $t \in T_L \rightarrow t$ ho elementi?
- VARIABILI DIPENDENTI:
  - $N_t \in \mathbb{Z}$, $k=1,\dots,K$, $t \in T_L \rightarrow$ numero di punti in t?
  - $N_{kt} \in \mathbb{Z}$, $k=1 \dots K$, $t \in T_L \rightarrow$ # punti con etichetta k?

STRUTTURA DECISION TREE → UNIVARIATO

l'unica cosa che sappiamo è che avremo relazioni del tipo:

- $ax \geq b$ → avremo bisogno di $a_t \in \mathbb{R}^p$, $b_t \in \mathbb{R}$ $\forall t \in T_{BRANCH}$
- per imporre un test univariato:

$$a_t \in \{0,1\}^p \quad e \quad \sum_{j=1}^{p} a_{tj} = 1$$

- per modellare la non-completezza allora:

$$d_t = \begin{cases} 1 & \text{se effettuo un branch su } t \\ 0 & \text{altrimenti} \end{cases}$$

da cui ricaviamo: $\boxed{\text{VINCOLI}}$

1. $\sum_{j=1}^{p} a_{tj} = d_t$
2. $0 \leq b_t \leq d_t$ $\quad \forall t \in T_{BRANCH}$
3. $d_t \leq d_{p(t)}$ $\quad \forall t \in T \setminus \{1\}$

ALLOCAZIONE PUNTI ALLE FOGLIE

dobbiamo introdurre altre variabili:

- $z_{it} = \begin{cases} 1 & \text{se il punto } x_i \in \text{alla foglia } t \\ 0 & \text{altrimenti} \end{cases}$

ovvero: → $z_{it} \in \{0,1\}$ per $i=1 \dots n$ e $t \in T_{LEAF}$

- $l_t = \begin{cases} 1 & \text{se la foglia } t \text{ ha assegnato qualche } x_i \\ 0 & \text{altrimenti} \end{cases}$

ovvero: → $l_t \in \{0,1\}$ per $t \in T_{LEAF}$

da cui ricaviamo i $\boxed{\text{VINCOLI}}$:

4. $z_{it} \leq l_t$ per $t \in T_{LEAF}$
5. $\sum_{t \in T_L} z_{it} = 1$ per $i=1 \dots n$ → ogni punto $\in$ TS deve $\in$ a una foglia
6. $\sum_{i=1}^{n} z_{it} \geq N_{min}\, l_t$ per $t \in T_L$ → N_{min}: numero minimo di punti da assegnare a una foglia

• possiamo anche aggiungere vincoli relativi al percorso RADICE-FOGLIA, ovvero ogni x_i deve rispettare tutti i vincoli che precedono la sua foglia.

$$\begin{cases} a_m^T x_i \leq b_m + M_1(1-z_{it}) \\ a_m^T x_i \geq b_m + M_2(1-z_{it}) \end{cases}$$

$\rightarrow$ NON LINEARE

linearizzo

$$a_m^T x_i + \underset{\text{piccola}}{\underline{\varepsilon}} < b_m + M_1(1-z_{it})$$

allora per $\forall$ features:

$$\varepsilon_j = \min\{|x_{ij} - x_{hj}| : i,h = 1,\dots,n\}$$

e definendo ε come il vettore:

$$\varepsilon^T = (\varepsilon_1 \cdots \varepsilon_p) \longrightarrow a_m^T(x_i + \varepsilon) \leq b_m + M_1(1-z_{it})$$

in cui nell'UNIVARIATO:

$$a_m^j = \begin{cases} 1 & j = J \\ 0 & j \in \{1 \dots p\} \setminus \{J\} \end{cases}$$

FUNZIONE OBIETTIVO

vogliamo misurare:

· COMPLESSITÀ (dimensione DT)

· MISSCLASSIFICATION_ERROR

allora ipotizzando di avere una struttura dati in cui:

· RIGHE: dati TS

· COLONNE: le classi $\rightarrow \{Y\}_{\substack{i=1,\dots,n \\ k=1,\dots,K}}$ ovvero $Y_{ik} = \begin{cases} 1 & \text{se } k\,y \\ -1 \end{cases}$

possiamo introdurre altre variabili:

• numero di punti nella foglia:

$$N_t = \sum_{i=1}^{n} z_{it} \quad \text{per } t \in T_L$$

• numero di punti nella foglia con etichetta k:

$$N_{kt} = \frac{1}{2}\sum_{i=1}^{n}(1+Y_{ik})\, z_{it} \quad \text{per } t \in T_L,\ k = 1 \dots K$$

• variabile che identifica l'etichetta più popolare nella foglia:

$$C_t = \underset{k=1,\dots,K}{\arg\max}\{N_{kt}\}$$

che ci portano a definire i $\boxed{\text{VINCOLI}}$:

(7) $\sum_{i=1}^{K} C_{it} = \ell_t$ per $t \in T_L$ (3)

e quindi: poi il costo di missclassification:

- $L_t = N_t - \max_{i=1 \dots K} \{N_{kt}\} \rightarrow$ va linearizzato

$\downarrow$

$$L_t \geq N_t - N_{kt} - M(1 - C_{kt})$$
$$L_t \leq N_t - N_{kt} + M(1 - C_{kt})$$

siccome $L_t \leq n$ allora:

$M = n \rightarrow$ ok

e definendo poi:

- errore complessivo (MISSCLASSIFICATION)

$$\ell = \sum_{t \in T_B} L_t$$

- COMPLESSITÀ del DT:

$$C = \sum_{t \in T_B} d_t$$

allora la $\boxed{\text{FUNZIONE OBIETTIVO}}$: $\lambda = \frac{1}{\hat{L}}$ = BASELINE ACCURACY

- $\min \left(\lambda \cdot \ell + \alpha \cdot C \right)$ dove $\begin{cases} \lambda \in (0,1) \\ \alpha \in (0,1) \end{cases}$

Per migliorare le PRESTAZIONI possiamo utilizzare un WARM START ovvero si cerca una soluzione valida per alberi a profondità $< D$ ~~[illegible]~~ facendo pruning sull'albero.

Bisogna quindi a questo punto scegliere:

- PROFONDITÀ MASSIMA D_{max}
- PARAMETRO DI COMPLESSITÀ α

• scelta di D:

si sceglie D e si incomincia a costruire da $D=0$ fino a D_{max} scelto

• scelta di α:

l'idea sarà, anziché usare i $\alpha \sum_{t \in T_B} d_t$ nella funzione obiettivo, lo portiamo nei vincoli.

VINCOLO

$$\sum_{t \in T_B} d_t = C \quad \text{dove } C = 1, \dots, C_{max} = 2^{D-1} = |T_B|$$

↳ max # split

DECISION TREE → MULTIVARIATO

partiamo sempre dalla relazione: 5

$a_t^T x_i \geq b_t$

in cui dobbiamo rilassare però il vincolo su a:

- $a \in [-1,1]^p$

allora bisogna linearizzare il vincolo di coerenza tra a e d_t:

- $d_t \in \{0,1\}$ → $\underline{\sum_{i=1}^{p} a_{it} = d_t}$

no linearizzato

$$\downarrow$$

$$\sum_{i=1}^{p} |a_{it}| \leq d_t$$

quindi i nuovi $\boxed{\text{vincoli}}$ saranno:

1. $\hat{a}_{it} \geq a_{it}$
2. $\hat{a}_{it} \geq -a_{it}$

→ $\hat{a}_{it} \geq |a_{it}|$ dove $\hat{a}_{it}$ = VARIABILE AUSILIARIA

3. $\sum_{i=1}^{p} \hat{a}_{it} \leq d_t$ per $t \in T_B$

mentre il vincolo su b_t diventa:

4. $-d_t \leq b_t \leq d_t$ $t \in T_B$

introduciamo poi nuove variabili:

- $s_{jt} = \begin{cases} 1 & \text{se l'attributo } j \text{ è usato nello split di } t \\ 0 & \text{altrimenti} \end{cases}$ (NODO ↓ t)

dove: → $s_{jt} \in \{0,1\}$ per $j=1,\dots,p$ e $t \in T_B$

che devono essere coerenti con le a_{it}

5. $-s_{it} \leq a_{it} \leq s_{it}$ per $i=1,\dots,p$, $t \in T_B$

possiamo poi aggiungere $\boxed{\text{VINCOLI ADDITIVI}}$ ridondanti:

6 7

- $s_{it} \le d_t$
- $\sum_{i=1}^{p} s_{it} \ge d_t$

e nella <u>F. OBIETTIVO</u> possiamo scrivere:

$$\alpha \sum_{t \in T_b} \sum_{j=1}^{p} s_{it} \rightarrow \text{termine complessità.}$$

CLASSIFICAZIONE ROBUSTA

ci aiuta nel mitigare ai rumori e incertezze che caratterizzano i TS.

Possiamo quindi scrivere:

$$\max C(x,u) : g(x,u) \leq 0^m ,\ x \in X$$

che in ottimizzazione robusta diventa

$$\max \left(\min \{ C(x,u) : g(u,x) \leq 0^m \} \right)$$

mettendoci nella situazione più brutta/pessimistica in cui la variazione su u tende a minimizzare quello che vogliamo massimizzare.

- Possiamo definire l'insieme di incertezza sulle features U_x (uncertain-set) come:
  - $x_i \in \mathbb{R}^p \rightarrow x_i + \Delta x_i$ con $\Delta x_i \in \mathbb{R}^p$
  - $\Delta x = \{\Delta x_1, \dots, \Delta x_n\}$

allora:

$$U_x = \{ \Delta x \in \mathbb{R}^{p \times n} : \| \Delta x_i \| \leq \underset{\substack{\uparrow \\ \text{parametro} \\ \text{magnitudo deciso}}}{\rho} ,\ i = 1 \dots n \}$$

Possiamo quindi <u>ROBUSTIFICARE OCT</u>.

definendo con:

- K = # nodi
- D = profondità albero

allora definiamo le seguenti variabili:

- $d_K = \begin{cases} 1 & \text{se non spitto sul nodo} \\ 0 & \text{altrimenti} \end{cases}$

dove: $\rightarrow d_K \in [0,1]$

- $z_{iK} \in \{0,1\} \rightarrow$ posso assegnare i, K

quindi i vincoli diventano: 9

① $d_k = 1$ dove $k = \lceil K/2 \rceil, \dots, k$ = foglie

② $d_k \geq d_u$ se ~~non~~ splitto allora ho già splittato

③ $d_k + \sum_{j=1}^{p} a_{kj} = 1$

④ $\sum_{k=1}^{K} z_{ik} = 1$ per $i = 1, \dots, n$

⑤ $z_{ik} \leq d_k$ (assegno i ad una foglia)

⑥ $z_{ik} \leq 1 - d_k$

introducendo poi:

- $c_k \in \{1, 0\}$ se $c_k = 1 \rightarrow$ nodo foglia

⑦ $\sum_{i=1}^{n} z_{ik} \geq N_{min} \cdot c_k$

⑧ $c_k \geq d_k - \sum_{u \in A(k)} d_u$

⑨ $a_u^T x_i + \varepsilon \leq b_u + M(1 - z_{ik})$ $\forall i = 1, \dots, n, \forall k \in T, \forall u \in A^L(k)$

⑩ $a_u^T x_i \geq b_u - M(1 - z_{ik})$ $''$, $''$, $\forall u \in A^R(k)$

Modello per la $\boxed{\text{FUNZIONE OBIETTIVO}}$:

- introduciamo:
  - 2 classi di etichette $y_i \in \{-1, 1\}$
  - $\underline{q_k, h_k}, \underline{f_k} \in \mathbb{Z}$, $w_k \in \{0, 1\}$

punti assegnati a k con etichetta → punti mal classificati

allora avremo $\boxed{\text{VINCOLI}}$:

⑪ $q_k = \sum_{i=1}^{m} \frac{1}{2}(1 - y_i) z_{ik}$

⑫ $h_k = \sum_{i=1}^{n} \frac{1}{2}(1 + y_i) z_{ik}$

⑬ $D - f_k \leq q_k + M[w_k + (1 - c_k)]$

allora possiamo Robustificare:

$\boxed{\text{VINCOLI ROBUSTIFICATI}}$

(9') $a_u^T x_i + P + E \le b_u + M(1 - z_{ik}) \quad u \in A^L(k)$

(10') $a_u^T x_i - P \ge b_u - M(1 - z_{ik}) \quad u \in A^R(k)$

(11') $g_k = \sum_{i=1}^{n} \frac{1}{2} \left(1 - y_i(1 - 2\Delta y_i)\right) z_{ik}$

(12') $h_k = \sum_{i=1}^{n} \frac{1}{2} \left(1 + y_i(1 - 2\Delta y_i)\right) z_{ik}$

allora lavorando sulla (11') possiamo ricavare la f. obiettivo, ottenendo:

$$\begin{cases} q^k = \min \sum_{i=1}^{n} (y_i z_{ik}) \cdot \Delta y_i \\ \sum_{i=1}^{n} \Delta y_i \le r \\ 0 \le \Delta y_i \le 1 \quad \text{per } i = 1 \dots n \end{cases}$$

di cui possiamo anche ricavare il duale:

$\mu_k \le 0$

$v_{ik} \le 0$

$$\longrightarrow \begin{cases} \max \left(\sum_{i=1}^{n} v_{ik}\right) + r\mu_k \\ v_{ik} + \mu_k \le y_i z_{ik} \\ \mu_k \le 0, \; v_k \le 0^n \end{cases}$$

ottenendo quindi da una qualunque soluzione ammissibile:

$$\left(\sum_{i=1}^{n} v_i\right) + \mu_k r \le q^k \quad \text{(per dualità debole)}$$

e quindi robustificato l'ultimo vincolo:

(13) $\sum_{i=1}^{n} \frac{1}{2}(1 - y_i) z_{ik} + \left(\sum_{i=1}^{n} v_{ik}\right) + r\mu_k \ge f_k M \left(w_k + (1 - c_k)\right)$

$v_{ik} + \mu_k \le y_i z_{ik}$

$\mu_k \le 0$

$v_{ik} \le 0$