1 主問題

1.1 所与の条件

特徴空間 (距離空間)	$(\mathcal{F}, d) \quad d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$
ラベル空間	$\mathcal{L}$
学習済み分類器	$f : \mathcal{F} \to \mathcal{L}$
目的データ	$x^* \in \mathcal{F}$
深さ制約	$D_{\max} \in \mathbb{N}$
精度制約	$A_{\min} \in [0, 1]$

1.2 決定木

決定木	$t : \mathcal{F} \to \mathcal{L}$
木の深さ	$D(t) : \mathcal{T}_{\mathcal{F}} \to \mathbb{N}$
仮説空間	$\mathcal{T}_{\mathcal{F}}$ — 特徴空間 $\mathcal{F}$ において, 可能な決定木の集合 $\mathcal{T}_{\mathcal{F}}(D_{\max}) = \{t \in \mathcal{T}_{\mathcal{F}} \mid D(t) \leq D_{\max}\}$

1.3 変数

近傍半径	$r \in \mathbb{R}_+$

1.4 関数など

近傍	$V_{x^*}(r) = \{x \in \mathcal{F} \mid d(x, x^*) \leq r\}$
ノイズ集合	$\mathrm{noise}(r) : \mathbb{R}_+ \to 2^F \quad \forall V \in 2^{\mathcal{F}}\ ;\ \mathrm{noise}(r) \subseteq V_{x^*}(r) \land \mathrm{noise}(r)$ is finite.

近似精度

$$A(t, r) = \frac{1}{|\mathrm{noise}(r)|} \sum_{x \in \mathrm{noise}(r)} \mathbb{I}(t(x) = f(x))$$

1.5 問題

$\exists t \in \mathcal{T}_{\mathcal{F}}(D_{\max})\ ;\ A(t, r) \geq A_{\min}$ を満足する最大の近傍半径 $r \in \mathbb{R}_+$ を求める.

2　固定されたデータセットの場合

2.1　所与の条件

特徴空間　$\mathcal{F}$

ラベル空間　$\mathcal{L}$

データセットのサイズ　$N \in \mathbb{N}$

データセット　$X = \{x_i \in \mathcal{F}\}_{i=1}^N, \quad Y = \{y_i \in \mathcal{L}\}_{i=1}^N$

目的データ　$x^* \in X$

深さ制約　$D_{\max} \in \mathbb{N}$

精度制約　$A_{\min} \in [0, 1]$

2.2　決定木

決定木　$t : \mathcal{F} \to \mathcal{L}$

木の深さ　$D(t) : \mathcal{T}_{\mathcal{F}} \to \mathbb{N}$

仮説空間　$\mathcal{T}_{\mathcal{F}}$ — 特徴空間 $\mathcal{F}$ において, 可能な決定木の集合

$\mathcal{T}_{\mathcal{F}}(D_{\max}) = \{t \in \mathcal{T}_{\mathcal{F}} \mid D(t) \leq D_{\max}\}$

2.3　関数など

近似精度

$$A_{X,Y}(t) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(y_i = t(x_i))$$

2.4　問題

$A_{X,Y}(t) \geq A_{\min}$ を満足する $t \in \mathcal{T}_{\mathcal{F}}(D_{\max})$ が存在するか否かを判定する.