第14章　モデルの結合

Pattern
Recognition
and
Machine
Learning

修士２年
山川佳洋

BehaviorInNetworks
都市生活学・ネットワーク行動学研究室

14章の目次

- モデルの結合
- ベイズモデル平均化
- コミッティ
- ブースティング
- 指数誤差の最小化
- ブースティングのための誤差関数

- 木構造モデル
- 条件付き混合モデル
- 線形回帰モデルの混合
- ロジスティックモデルの混合
- 混合エキスパートモデル

概要

- コミッティ
  L個の異なるモデルを訓練した後に，各モデルで得られた予測の平均値を予測値として用いる
  →代表的なものにブースティング
- 決定木
  予測に用いる1つのモデルを入力変数の関数として選択するもの
  →応用したものに混合エキスパートモデル

コミッティ

- L個の異なるモデルを訓練した後に，各モデルで得られた予測の平均値を予測値として用いる
- 各モデル間には変化が必要
→バギングの利用

M 個のブートストラップデータ集合を生成し，それらデータ集合を用いて個々に独立な M 個の予測モデル $y_m(\mathbf{x})$ のコピーを訓練する.

$$y_{COM}(\mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} y_m(\mathbf{x}) \quad (14.7)$$

ブースティング

- コミッティとの違い
  →データを逐次的に訓練
- 複数の「ベース」分類器を結合する
  →いずれのベース分類器より高性能のコミッティ
  →ベース分類器は弱学習器と呼ばれる
- 代表的なものにAdaBoost (Freund and Schapire, 1996)
- もとは分類問題→回帰問題にも拡張
- 各ベース分類器の訓練→重み付けられたデータ集合
- 重み係数は以前の学習の分類器の性能による

AdaBoost アルゴリズム

1. n=1,…,N のデータの重み係数$\{\omega_m\}$を $\omega_n^{(1)}=1/N$ に初期化する
2. m=1,…,M について以下を繰り返す

(a)分類器 $y_m(\mathbf{x})$ を次の重み付けされた誤差関数を最小化するように訓練データにフィットさせる

$$J_m=\sum_{n=1}^{N}\omega_n^{(m)}I\left(y_m(\mathbf{x}_n)\neq t_n\right) \quad (14.15)$$

(b)次の値(誤差率の尺度)を計算する

$$\varepsilon=\frac{\sum_{n=1}^{N}\omega_n^{(m)}I\left(y_m(\mathbf{x}_n)\neq t_n\right)}{\sum_{n=1}^{N}\omega_n^{(m)}} \quad (14.16)$$

これを用いて次の量(重み係数)を求める

$$\alpha_m=\ln\left\{\frac{1-\varepsilon_m}{\varepsilon_m}\right\} \quad (14.17)$$

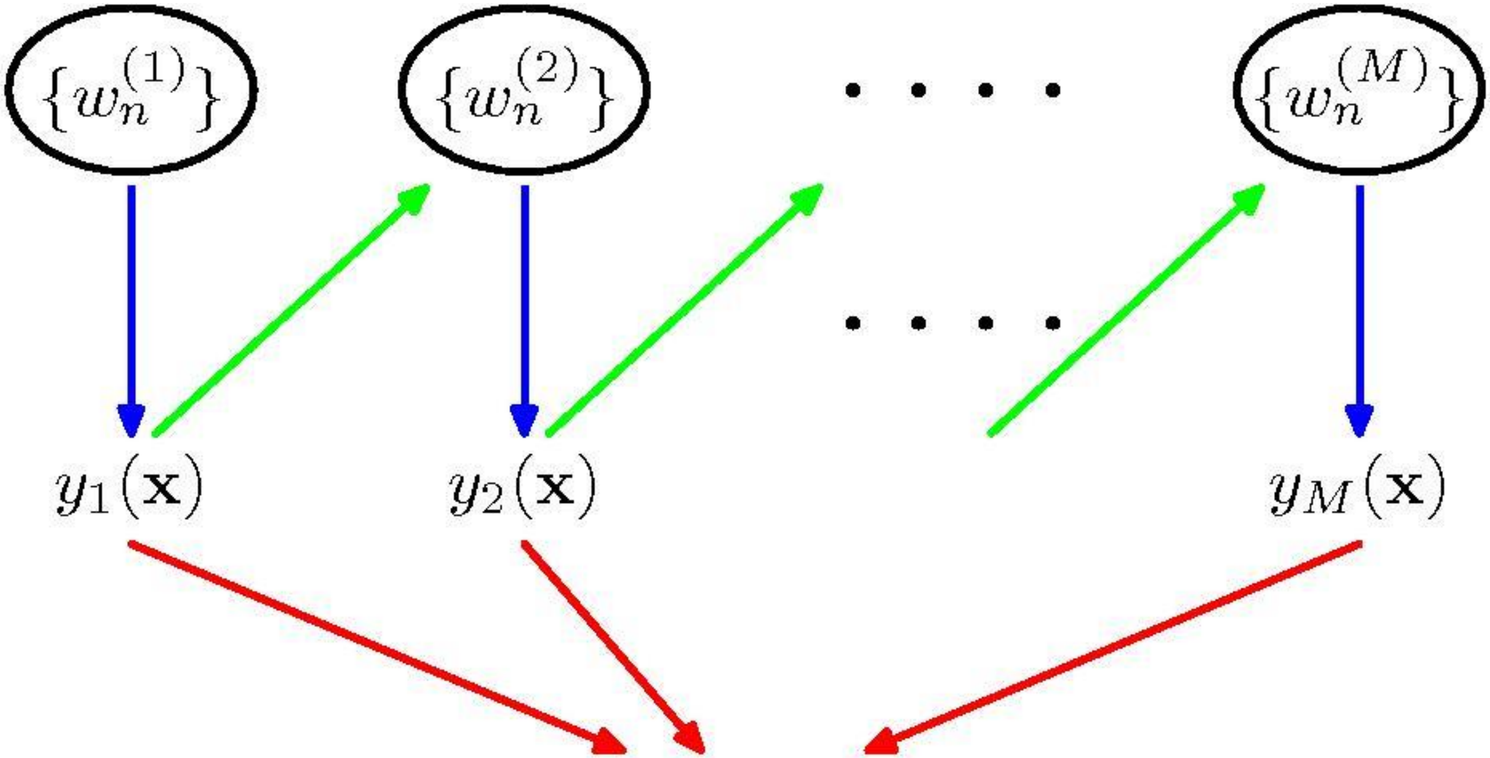
(c)データ点の重み係数を以下の式で更新する

$$\omega_n^{(m+1)}=\omega_n^{(m)}\exp\left\{\alpha_m I\left(y_m(\mathbf{x}_n)\neq t_n\right)\right\} \quad (14.18)$$

3. 以下の式で，最終モデルの予測をする

$$Y_M(\mathbf{x})=sign\left(\sum_{m=1}^{M}\alpha_m y_m(\mathbf{x})\right) \quad (14.19)$$

AdaBoost アルゴリズム



$$Y_M(\mathbf{x}) = \mathrm{sign}\left(\sum_{m}^{M} \alpha_m y_m(\mathbf{x})\right)$$

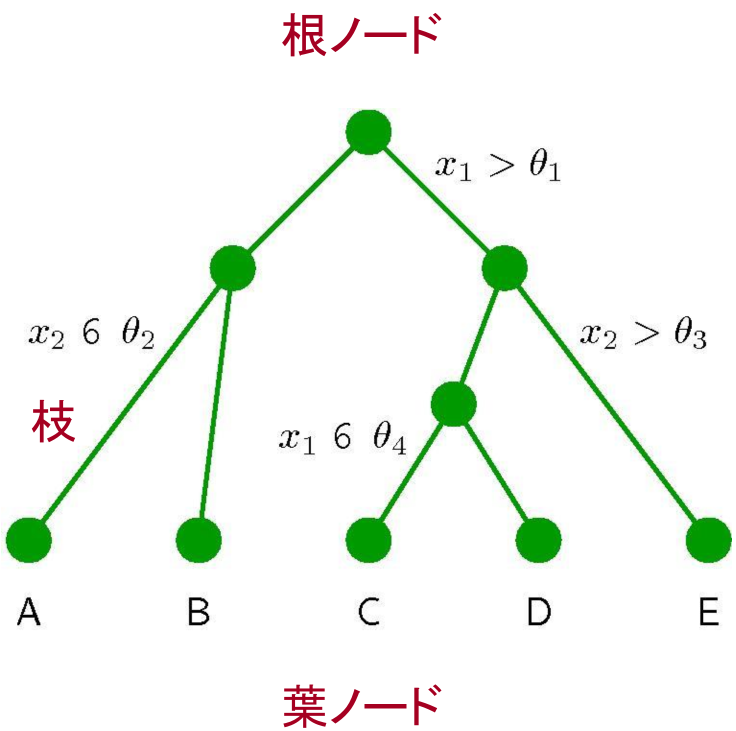
決定木

- 木構造に沿った一連の二値選択として記述
- 個々には非常に簡単なモデルを用いる
- 分類問題，回帰問題，いずれにも適用

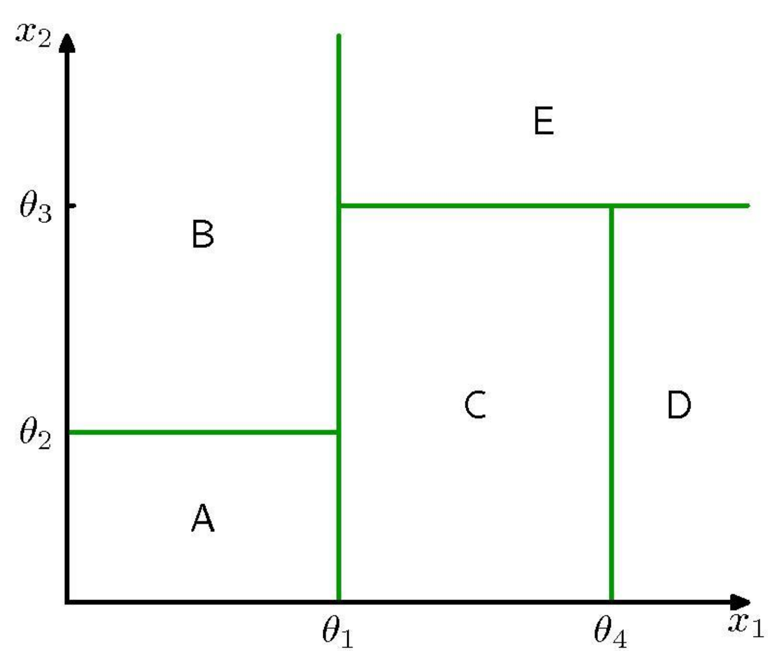
決定木　例１(PRMLより)

決定木

木構造の予測モデル
入力空間を多次元の矩形領域に区分する

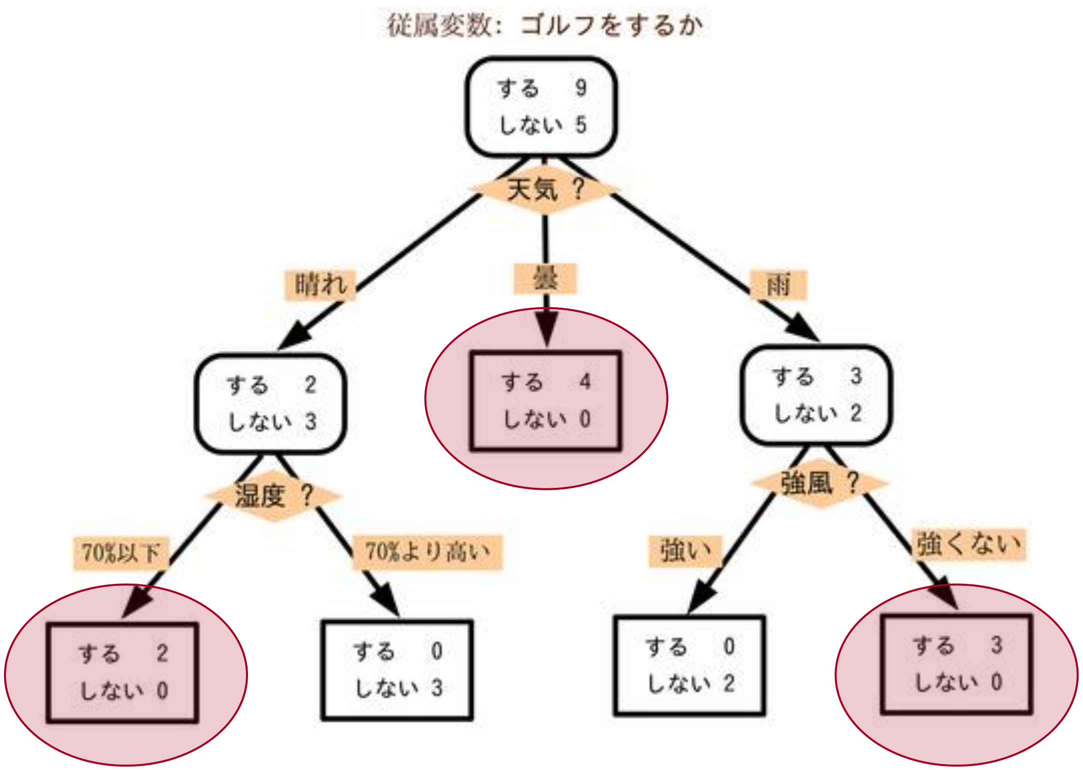


二次元入力空間



決定木　例２(wikipediaより)

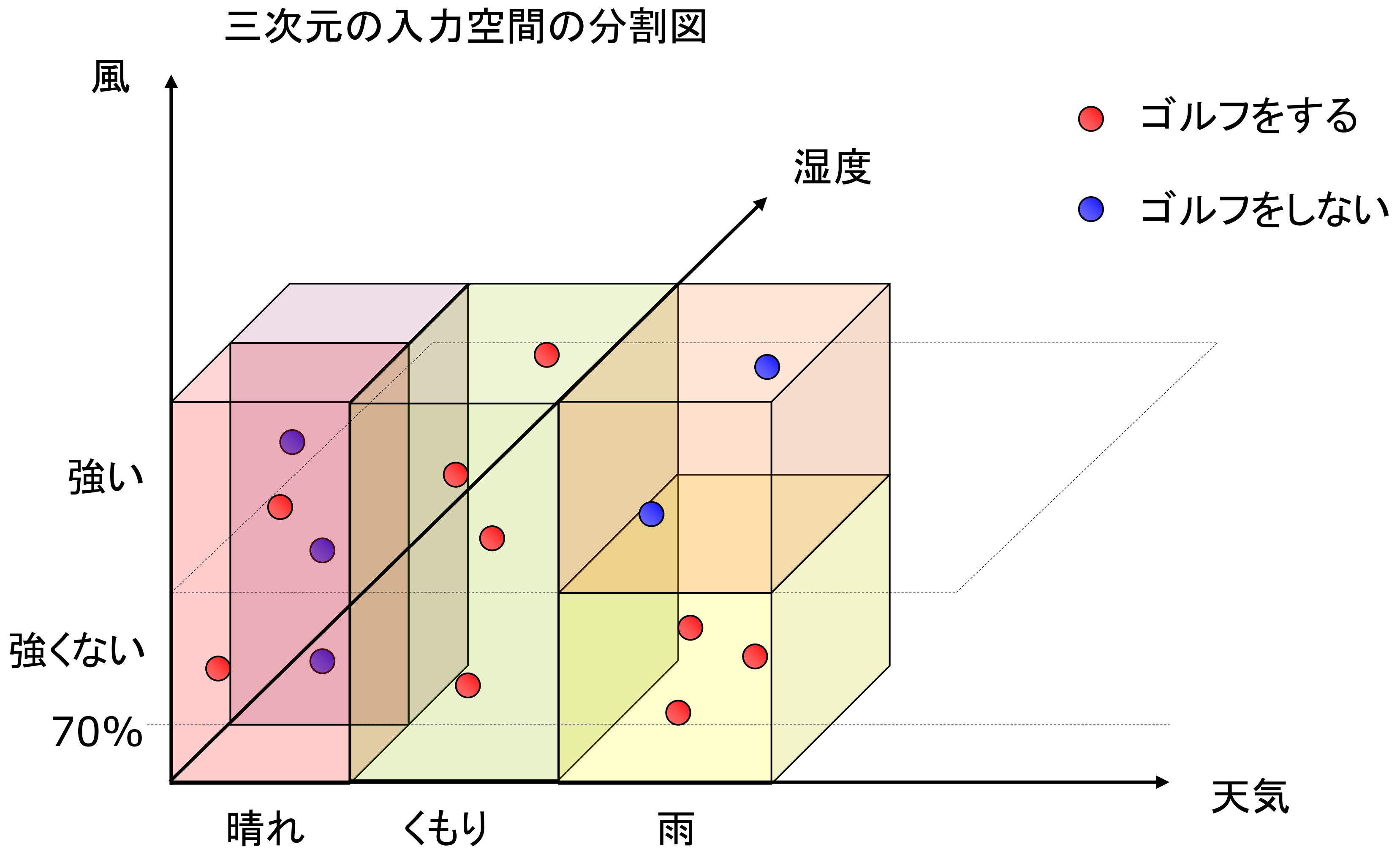
ゴルフ場の経営者が従業員の勤務体制を最適化する



独立変数				従属変数
天気	気温 (度)	湿度 (%)	風が強いか	ゴルフをするか
晴れ	22	95	強くない	しない
晴れ	21	70	強くない	する
晴れ	24	70	強い	する
晴れ	29	85	強くない	しない
晴れ	27	90	強い	しない
曇	27	75	強くない	する
曇	28	78	強くない	する
曇	22	90	強い	する
曇	18	65	強い	する
雨	22	80	強い	しない
雨	21	96	強くない	する
雨	20	80	強くない	する
雨	18	70	強い	しない
雨	24	80	強くない	する

独立変数				従属変数
天気	気温 (度)	湿度 (%)	風が強いか	ゴルフをするか
曇	18	65	強い	する
雨	18	70	強い	しない
雨	20	80	強くない	する
晴れ	21	70	強くない	する
雨	21	96	強くない	する
雨	22	80	強い	しない
曇	22	90	強い	する
晴れ	22	95	強くない	しない
晴れ	24	70	強い	する
雨	24	80	強くない	する
晴れ	27	90	強い	しない
曇	27	75	強くない	する
曇	28	78	強くない	する
晴れ	29	85	強くない	しない

決定木　例２イメージ図



決定木！

目標変数を予測するためのモデルは各領域に個別に存在

回帰問題では領域ごとに単純に定数値を予測
→家賃の見積もり(駅からの距離，広さ，築年数など)

分類問題では各領域に特定のクラスを割り当てる
→医療診断(体温，血圧など)

訓練集合からの学習

各ノードにおいて分割規準として利用する入力変数を選択肢し
閾値θiを決めることで木構造を決定する

領域ごとに予測する変数の値を決定する

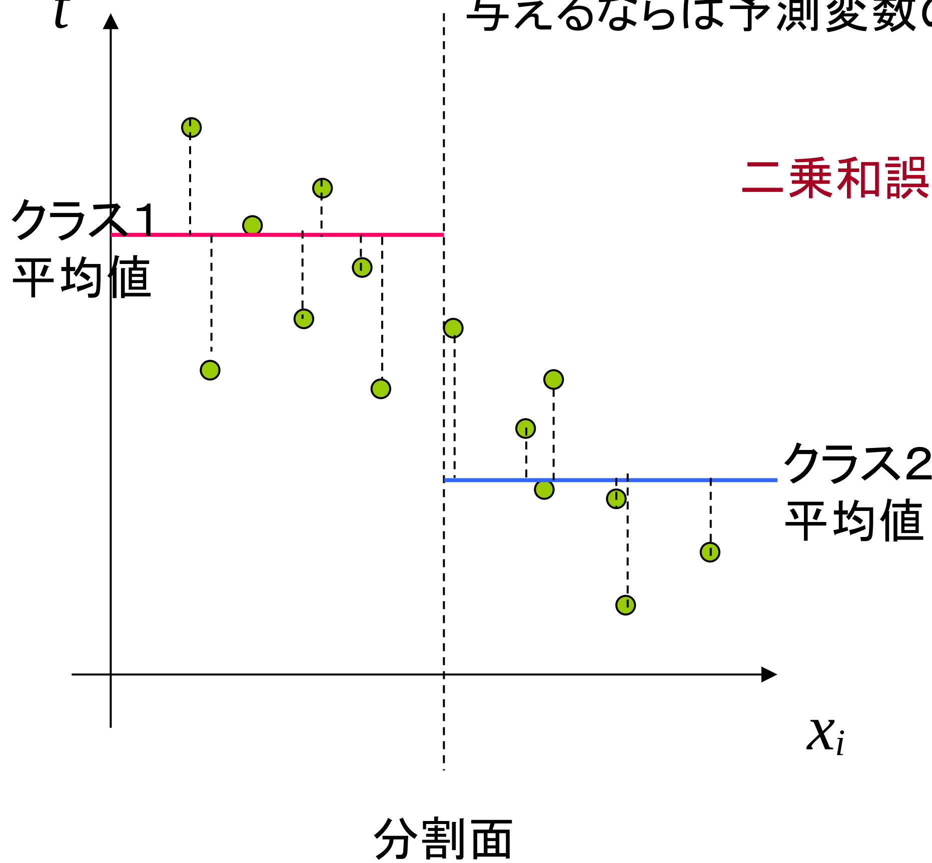
D次元ベクトル $\mathbf{x} = (x_1, ..., x_D)^T$ から一次元の目標変数 t を予測する

訓練データ

連続ラベル $\{t_1, ..., t_N\}$ を伴う入力ベクトル $\{x_1, ..., x_N\}$

回帰モデルの分割方法

入力空間の分割を二乗誤差を最小にするように
与えるならば予測変数の最適値は領域内のデータ点の平均値となる



二乗和誤差が最小となるように分割を行う

ノードの追加を終わらせる条件　　T : 葉

$$y_\tau = \frac{1}{N_\tau} \sum_{x_n \in R_\tau} t_n \qquad (14.29)$$

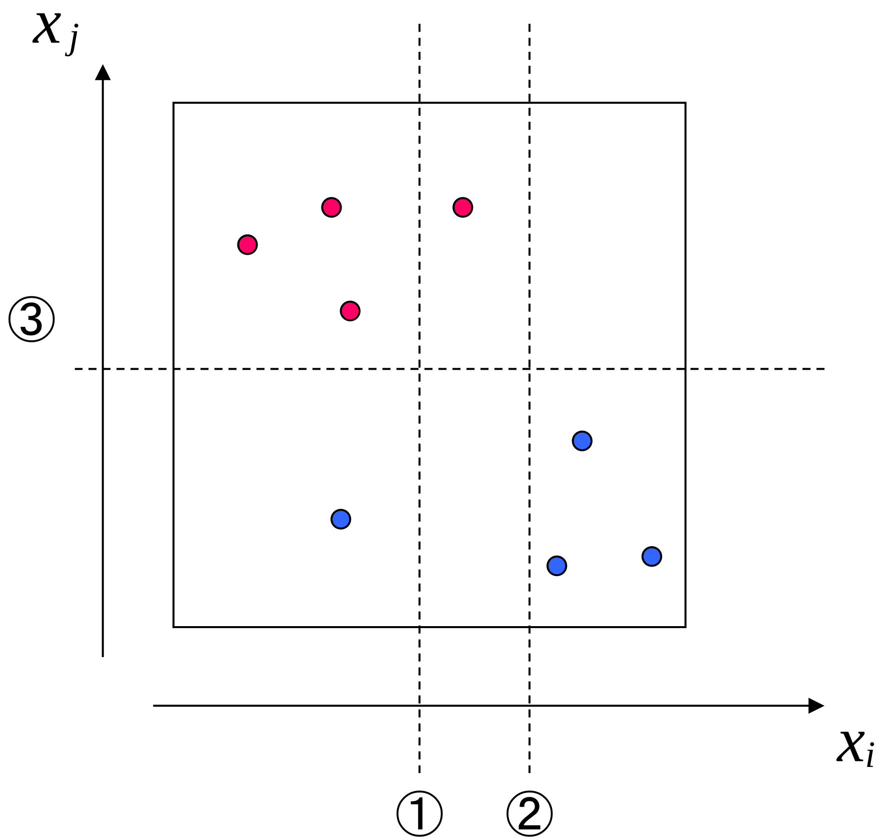
$$Q_\tau(T) = \sum_{x_n \in R_\tau} \{t_n - y_\tau\}^2 \qquad (14.30)$$

$$C(T) = \sum_{\tau=1}^{|T|} Q_\tau(T) + \lambda |T| \qquad (14.31)$$

分類問題の分割方法

ジニ係数

$$Q_\tau(T) = \sum_{k=1}^{K} p_{\tau k}(1 - p_{\tau k}) \quad (14.33)$$



最小となるように分割

2クラス分類(赤と青)

① $\frac{3}{4} * \frac{1}{4} + \frac{1}{4} * \frac{3}{4} = \frac{3}{8}$

② $\frac{4}{5} * \frac{1}{5} + \frac{0}{3} * \frac{3}{3} = \frac{4}{25}$

③ $\frac{4}{4} * \frac{0}{4} + \frac{0}{4} * \frac{4}{4} = 0$

③<②<①

決定木！！

- 人における可読性が木モデルの強み
- データ集合の細部に非常に敏感
  →データのわずかな違いから結果が大きく変わることも
- 分割が特徴空間の軸に沿わせているため準最適となる
- 回帰問題で予測が分離境界において不連続
- 入力空間分割がハードな分割
  →確率的な枠組みの導入でソフトに
  →混合エキスパートモデル

$$p(t|\mathbf{x}) = \sum_{k=1}^{K} \pi_k(\mathbf{x}) p(t|\mathbf{x}, k) \quad (14.53)$$

$$\pi_k(\mathbf{x}) = p(k|\mathbf{x})$$