機械学習 演習問題 解答例

立命館大学 情報理工学部

村上 陽平、福森 隆寛

Beyond Borders

演習問題の解答例

- これは講義中に実施した演習問題の解答例です
  - あくまでも「解答例」ですので、他にも正しい解答があることを理解した上で、この資料を参考にしてください

演習問題1-1（5分間）

- ビッグデータを用いた機械学習の応用事例を最低１つ考えなさい。

例）翻訳システム

２ヶ国語で記述された文章データを大量に集めて、そのビッグデータから翻訳のルール（規則）を機械学習によって導き出す。

演習問題1-1（5分間）解答例

- ビッグデータを用いた機械学習の応用事例を最低１つ考えなさい。

感染症の拡大予想

もし

① とても優秀な機械学習の技術が存在する

② 誰が、いつ、どこで、どのような症状にかかったのか？

という情報が記録されたビッグデータを迅速に集められる

が実現していれば、新型コロナウイルス（COVID-19）の流行をもっと早く予想できたかも？

演習問題1-2（10分間）

- 以下の学習手法について応用事例を1つずつ考えなさい。
  - 教師あり学習（回帰）
  - 教師なし学習（モデル推定）
  - 半教師あり学習
  - 強化学習

※ 講義スライドに記載されていない応用事例を考えてください。

演習問題1-2 （10分間）解答例

- 以下の学習手法について応用事例を1つずつ考えなさい。
  - 教師あり学習（回帰）

株価(かぶか)の予測

過去の株価情報や経済状況(けいざいじょうきょう)をもとに、未来(みらい)の株価を予測する。
高精度(こうせいど)な予測システムを構築できれば、ビジネスチャンスに繋(つな)がるかも…

  - 教師なし学習（モデル推定）

開発商品の改善(かいぜん)点の抽出

商品に対する口(くち)コミ（人伝(ひとづて)で伝達(でんたつ)される評判(ひょうばん)や評価に関する情報）の文書(ぶんしょ)をクラスタリングして、典型的(てんけいてき)な不満(ふまん)や要望(ようぼう)を抽出する。

演習問題1-2 （10分間）解答例

- 以下の学習手法について応用事例を1つずつ考えなさい。
  - 半教師あり学習

口コミのPN判定（Positive-Negative判定）

ネットワーク上(じょう)の口コミのPN判定では、正解付きデータを収集するのは大変(たいへん)（収集できても少量(しょうりょう)）だが、クロープログラムを使えば、自動的に大量の教師なしデータを収集できる。ある製品(せいひん)に対する口コミのPN判定は、半教師あり学習が有効。

  - 強化学習

囲碁(いご)の最善(さいぜん)の一手(いって)（Alpha Goなど）

相手(あいて)に勝利(しょうり)するまでの過程(かてい)は様々(さまざま)です。強化学習は、盤面(ばんめん)という状態に応(おう)じて、将来(しょうらい)の得られる報酬が1番高い（勝利の可能性(かのうせい)が高い）一手を導(みちび)きます。

演習問題2-1（10分間）

- 以下5名分の身長(しんちょう)と体重(たいじゅう)のデータについて、身長と体重の平均値を0、標準偏差を1となるように標準化せよ。

標準化前

番号	身長 [cm]	体重 [kg]
1	160	54
2	166	58
3	168	60
4	172	62
5	184	66
平均		
標準偏差		

標準化後

番号	身長	体重
1		
2		
3		
4		
5		
平均	0	0
標準偏差	1	1

【ヒント】標準化の方法

$$\text{標準化後の値} = \frac{\text{もとの値} - \text{その次元の平均値}}{\text{その次元の標準偏差}}$$

演習問題2-1（10分間）解答例

- 以下5名分の身長と体重のデータについて、身長と体重の平均値を0、標準偏差を1となるように標準化せよ。

標準化前

番号	身長 [cm]	体重 [kg]
1	160	54
2	166	58
3	168	60
4	172	62
5	184	66
平均	170	60
標準偏差	8	4

標準化後（標準化後の特徴に単位なし）

番号	身長	体重
1	-1.25	-1.5
2	-0.5	-0.5
3	-0.25	0
4	0.25	0.5
5	1.75	1.5
平均	0	0
標準偏差	1	1

演習問題2-2（10分間）

- ある画像認識技術でバナナを検出することを考える。
  - 100個のリンゴ、30個のバナナ、70個のオレンジの画像データに対して、以下の識別結果が得られた。

1. 識別結果から混同行列を求めよ。
2. 混同行列から正解率、精度、再現率、F値を計算せよ。
また、この認識技術の性能について考察(こうさつ)せよ。

	識別結果	
	バナナ	バナナ以外
りんご	20	80
バナナ	15	15
オレンジ	10	60

	予測＋	予測－
正解＋		
正解－		

正解率

精度

再現率

F値

演習問題2-2（10分間）解答例

- ある画像認識技術でバナナを検出することを考える。
  - 100個のリンゴ、30個のバナナ、70個のオレンジの画像データに対して、以下の識別結果が得られた。

1. 識別結果から混同行列を求めよ。

	識別結果	
	バナナ	バナナ以外
りんご	20	80
バナナ	15	15
オレンジ	10	60

混同行列

	予測＋	予測－
正解＋	(TP) 15	(FN) 15
正解－	(FP) 20+10=30	(TN) 80+60=140

演習問題2-2（10分間）解答例

- ある画像認識技術でバナナを検出することを考える。
  - 100個のリンゴ、30個のバナナ、70個のオレンジの画像データに対して、以下の識別結果が得られた。

2. 混同行列から正解率、精度、再現率、F値を計算せよ。
また、この認識技術の性能について考察せよ。

混同行列

	予測＋	予測－
正解＋	(TP) 15	(FN) 15
正解－	(FP) 30	(TN) 140

$$\text{正解率} = \frac{\mathrm{TP+TN}}{\mathrm{TP+FN+FP+TN}} = \frac{15+140}{15+15+30+140} = 0.775$$

$$\text{精度} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} = \frac{15}{15+30} = \frac{1}{3}$$

→ バナナと識別した結果の内（うち）、33.3%しか正しくない

$$\text{再現率} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} = \frac{15}{15+15} = 0.5$$

→ バナナのデータを入力しても正しく識別できたものは50%

$$\text{F値} = 2 \times \frac{\text{精度} \times \text{再現率}}{\text{精度} + \text{再現率}} = 0.4$$

正解率が0.775（77.5%）なので、一見（いっけん）、良い結果のように見えるが、精度や再現率などの結果から、正確にバナナを検出できていない

演習問題3-1（10分間）

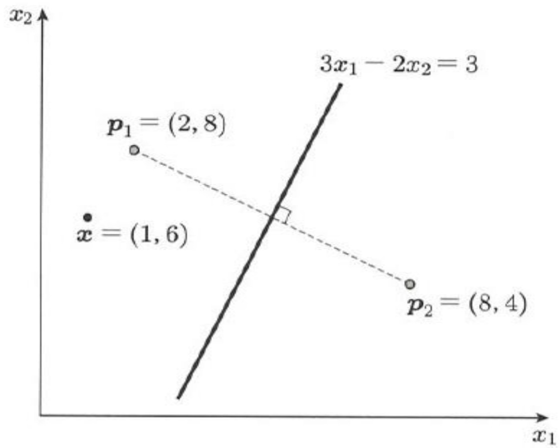
- 以下の2つのクラスのプロトタイプが与えられたとき、入力 $\boldsymbol{x} = (1,6)$ は最近傍決定則を用いるとどちらのクラスに識別されるか？
  - クラスω_1のプロトタイプ：$\boldsymbol{p}_1 = (2,8)$
  - クラスω_2のプロトタイプ：$\boldsymbol{p}_2 = (8,4)$

- このときの識別面の式を求めよ。

演習問題3-1（10分間）解答例

- $\boldsymbol{x}$と$\boldsymbol{p}_1, \boldsymbol{p}_2$ のユークリッド距離Dは
  - $D(\boldsymbol{x}, \boldsymbol{p}_1) = \sqrt{(1-2)^2 + (6-8)^2} = \sqrt{5}$
  - $D(\boldsymbol{x}, \boldsymbol{p}_2) = \sqrt{(1-8)^2 + (6-4)^2} = \sqrt{53}$
  - $D(\boldsymbol{x}, \boldsymbol{p}_1) < D(\boldsymbol{x}, \boldsymbol{p}_2)$ なので、入力$\boldsymbol{x}$はクラスω_1と判定

- 2つのプロトタイプの中点(5,6)を通り、直線$\boldsymbol{p}_1\boldsymbol{p}_2$と直交する直線(識別面)は $3x_1 - 2x_2 = 3$



演習問題3-2（10分間）

- 前スライドのパーセプトロンの学習規則を用いて以下の表のデータから識別関数$g(\boldsymbol{x})$を求めよ。
  - 重みの初期値：$w_0 = 0.2, w_1 = 0.3$
  - 学習係数：$\rho = 0.5$

クラス	x
ω_1	1.0
ω_1	0.5
ω_2	-0.2
ω_2	-1.3

演習問題3-2（10分間）解答例

- 与えられた初期値 $w_0 = 0.2, w_1 = 0.3$ に対応する識別関数 $g(\boldsymbol{x})$

$$g(\boldsymbol{x}) = \boldsymbol{w}_{(0)}^T \boldsymbol{x} = (w_0, w_1) \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = 0.3x_1 + 0.2$$

- この識別関数に
  - $x_0 = 1$
  - $x_1 = \{1.0, 0.5, -0.2, -1.3\}$（前スライドの表の$x$）

を代入して、クラスω_1のデータ$\{1.0, 0.5\}$に対して正の値、クラスω_2のデータ$\{-0.2, -1.3\}$に対して負の値を出力するように重み w_0 と w_1 を調整する

演習問題3-2（10分間）解答例

学習1巡目

クラス	x_0	x_1	w_0	w_1	$g(x)$	判定
ω_1	1	1.0	0.2	0.3	0.5	○
ω_1	1	0.5	0.2	0.3	0.35	○
ω_2	1	-0.2	0.2	0.3	0.14	誤判定（$\omega_2 \rightarrow \omega_1$）
ω_2	1	-1.3	-0.3	0.4	-0.82	○

（＊）

（＊）誤識別が生じた時点で、そのデータを使って重みを修正

$$\boldsymbol{w}' = \boldsymbol{w} - \rho \boldsymbol{x}$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix} - 0.5 \begin{pmatrix} 1.0 \\ -0.2 \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0.4 \end{pmatrix}$$

誤識別が発生しなくなるまで、同じことを繰り返す

演習問題3-2（10分間）解答例

学習2巡目

クラス	x_0	x_1	w_0	w_1	$g(x)$	判定
ω_1	1	1.0	−0.3	0.4	0.1	○
ω_1	1	0.5	−0.3	0.4	−0.1	誤判定（$\omega_1 \to \omega_2$）
ω_2	1	−0.2	0.2	0.65	0.07	誤判定（$\omega_2 \to \omega_1$）
ω_2	1	−1.3	−0.3	0.75	−1.275	○

（＊）

（＊＊）

（＊）　$$\boldsymbol{w}' = \boldsymbol{w} + \rho \boldsymbol{x}$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0.4 \end{pmatrix} + 0.5 \begin{pmatrix} 1.0 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.65 \end{pmatrix}$$

（＊＊）　$$\boldsymbol{w}' = \boldsymbol{w} - \rho \boldsymbol{x}$$

$$\begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.65 \end{pmatrix} - 0.5 \begin{pmatrix} 1.0 \\ -0.2 \end{pmatrix} = \begin{pmatrix} -0.3 \\ 0.75 \end{pmatrix}$$

誤識別が発生しなくなるまで、同じことを繰り返す

演習問題3-2（10分間）解答例

学習3巡目

クラス	x_0	x_1	w_0	w_1	$g(x)$	判定
ω_1	1	1.0	-0.3	0.75	0.45	○
ω_1	1	0.5	-0.3	0.75	0.075	○
ω_2	1	-0.2	-0.3	0.75	-0.45	○
ω_2	1	-1.3	-0.3	0.75	-1.275	○

全てのデータが正しく識別された

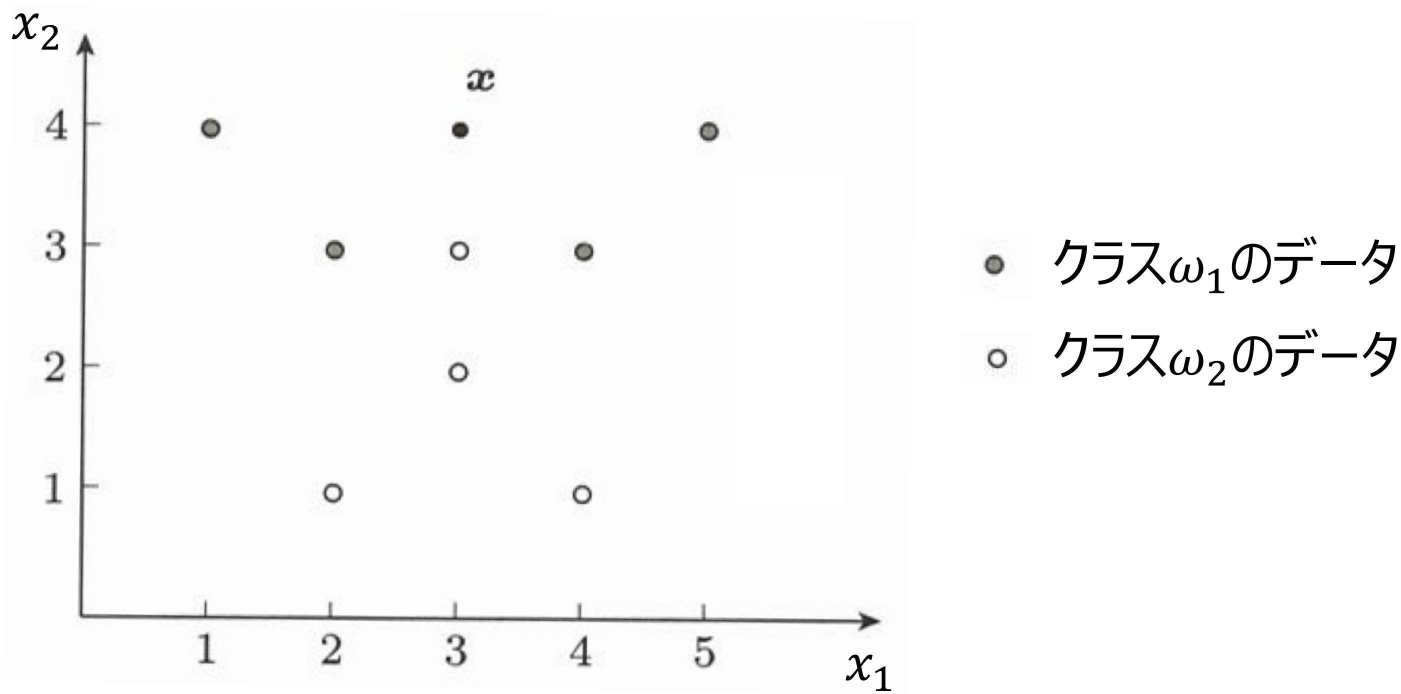
- 得られた識別関数は

$$g(\boldsymbol{x}) = (-0.3, 0.75)\begin{pmatrix} 1 \\ x_1 \end{pmatrix} = 0.75x_1 - 0.3$$

である。

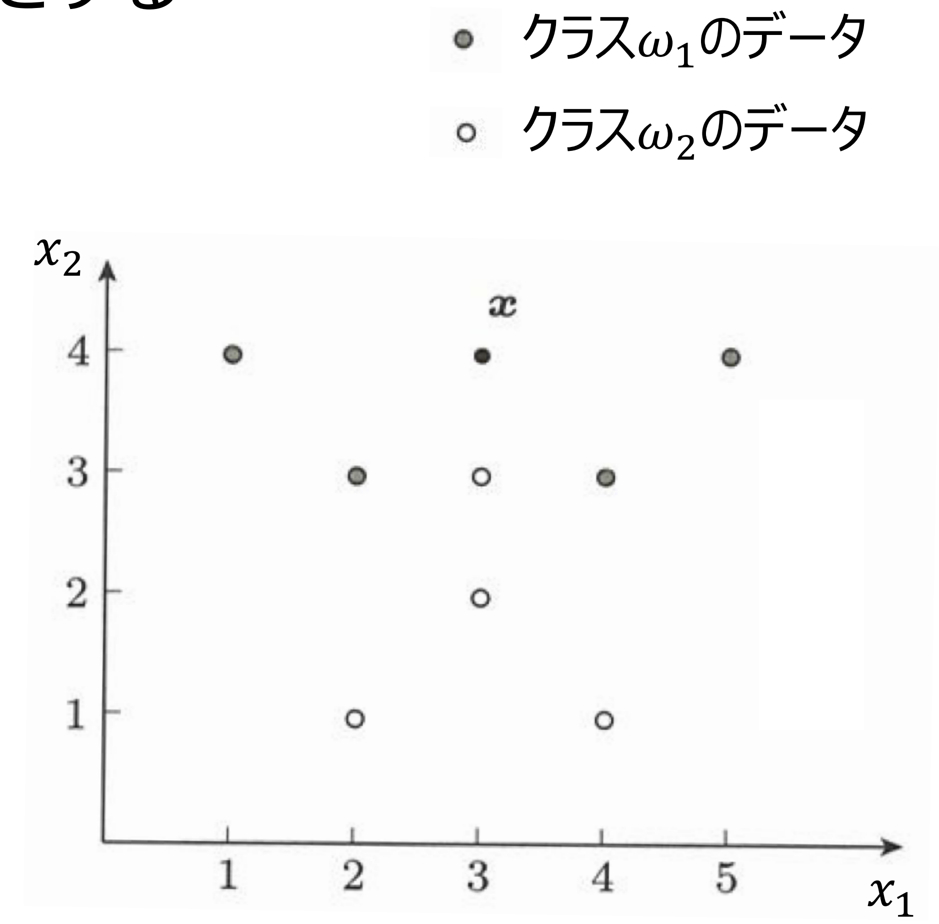
演習問題4-1（10分間）

- 下図に示す学習データを用いて、
  3-NN法によって入力$\boldsymbol{x} = (3,4)$を識別せよ
  - 識別結果の決め方は多数決とする



演習問題4-1（10分間）解答例

- 下図に示す学習データを用いて、
3-NN法によって入力$\boldsymbol{x} = (3,4)$を識別せよ
  - 識別結果の決め方は多数決とする



- $\boldsymbol{x}$に近いデータ（上位3個）
  1. クラスω_2のデータ (3,3)
  2. クラスω_1のデータ (2,3)
  3. クラスω_1のデータ (4,3)

クラスω_1のデータが2個、
クラスω_2のデータが1個なので
多数決より、$\boldsymbol{x}$の識別結果はクラスω_1

演習問題4-2（10分間）

- 次の文は、それぞれ「パーセプトロンの学習規則」と「Widrow-Hoffの学習規則」のどちらについて説明しているか？
  1. 識別関数の値と教師信号の二乗誤差の総和（そうわ）を最小化する
  2. 線形分離可能の場合でも、全ての学習パターンが正しく識別される重みが得られるとは限らない
  3. 線形分離不可能の場合は、学習は収束しない

演習問題4-2（10分間）解答例

- 次の文は、それぞれ「パーセプトロンの学習規則」と「Widrow-Hoffの学習規則」のどちらについて説明しているか？

1. 識別関数の値と教師信号の二乗誤差の総和を最小化する
   - 答え：Widrow-Hoffの学習規則
2. 線形分離可能の場合でも、全ての学習パターンが正しく識別される重みが得られるとは限らない
   - 答え：Widrow-Hoffの学習規則
3. 線形分離不可能の場合は、学習は収束しない
   - 答え：パーセプトロンの学習規則

演習問題4-2（10分間）解答例

- パーセプトロンの学習規則
  - 2クラスの場合、識別関数の正負と教師信号の正負が一致(いっち)するまで修正を繰り返す
  - 線形分離可能であれば、学習は収束し、全ての学習パターンが正しく識別される重みベクトルが得られる
  - 線形分離不可能の場合は、学習は収束しない

- Widrow-Hoff の学習規則
  - 識別関数の値と教師信号との二乗誤差の総和を最小化(さいしょうか)する
  - 線形分離不可能の場合でも収束する
  - 線形分離可能の場合でも、全ての学習パターンが正しく識別される重みが得られるとは限(かぎ)らない

演習問題5-1（10分間）

- 外見では区別のつかない箱Aと箱Bがある
  - 箱A：白玉が1個、黒玉が3個入っている
  - 箱B：白玉が4個、黒玉が1個入っている

1. 箱A、箱Bのいずれかから白玉を取り出す確率を求めよ。
   それぞれの箱の選び方は等確率とする。
2. 箱A、箱Bのいずれかから玉を取り出すと、白玉であった。
   この白玉が箱A、箱Bから取り出された確率をそれぞれ求めよ。
   それぞれの箱の選び方は等確率とする。
3. 箱A、箱Bの選ばれる確率が 9 : 1 であったとき、
   1. と 2. で求めた確率は、どのように変化するか？

演習問題5-1（10分間）解答例

1. 箱A、箱Bのいずれかから白玉を取り出す確率を求めよ。それぞれの箱の選び方は等確率とする。

- 「箱Aを選ぶ」かつ「箱Aから白玉が出る」の確率：$\frac{1}{8}$
  - 箱Aを選ぶ確率を$P(A)$、箱Aから白玉が出る確率を $P(白|A)$とすると
  $$P(A)P(白|A) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$
- 「箱Bを選ぶ」かつ「箱Bから白玉が出る」：$\frac{2}{5}$
  - 箱Bを選ぶ確率を$P(B)$、箱Bから白玉が出る確率を $P(白|B)$とすると
  $$P(B)P(白|B) = \frac{1}{2} \cdot \frac{4}{5} = \frac{2}{5}$$
- 白玉を取り出す可能性を全て加算（かさん）すれば良いので

$$P(白) = P(A)P(白|A) + P(B)P(白|B) = \frac{1}{8} + \frac{2}{5} = \mathbf{\frac{21}{40}}$$

演習問題5-1（10分間）解答例

2. 箱A、箱Bのいずれかから玉を取り出すと、白玉であった。この白玉が箱A、箱Bから取り出された確率をそれぞれ求めよ。それぞれの箱の選び方は等確率とする。

- 白玉が箱Aから取り出された確率：$P(A|白)$
- 白玉が箱Bから取り出された確率：$P(B|白)$
- ベイズの定理と1.の結果から

$$\boldsymbol{P(A|白)} = \frac{\mathrm{P}(白|A)P(A)}{P(白)} = \frac{\frac{1}{4}\times\frac{1}{2}}{\frac{21}{40}} = \boldsymbol{\frac{5}{21}}$$

$$\boldsymbol{P(B|白)} = \frac{\mathrm{P}(白|B)P(B)}{P(白)} = \frac{\frac{4}{5}\times\frac{1}{2}}{\frac{21}{40}} = \boldsymbol{\frac{16}{21}}$$

箱の選び方が等確率という条件で、白玉が出たときに、**箱Bを選択した**可能性が高いと考えられる

演習問題5-1（10分間）解答例

3. 箱A、箱Bの選ばれる確率が 9 : 1 であったとき、
   1. と 2. で求めた確率は、どのように変化するか？

■ 1. と 2. の計算式に $P(A)=\frac{9}{10}$、$P(B)=\frac{1}{10}$ を代入（だいにゅう）して再計算

$$P(白)=P(A)P(白|A)+P(B)P(白|B)=\frac{1}{4}\times\frac{9}{10}+\frac{4}{5}\times\frac{1}{10}=\mathbf{\frac{61}{200}}$$

$$\boldsymbol{P(A|白)}=\frac{\mathrm{P}(白|A)P(A)}{P(白)}=\frac{\frac{1}{4}\times\frac{9}{10}}{\frac{61}{200}}=\mathbf{\frac{45}{61}}$$

$$\boldsymbol{P(B|白)}=\frac{\mathrm{P}(白|B)P(B)}{P(白)}=\frac{\frac{4}{5}\times\frac{1}{10}}{\frac{61}{200}}=\mathbf{\frac{16}{61}}$$

箱A、箱Bの選ばれる確率が9:1であったという条件だと
白玉が出たときに、**箱Aを選択した**可能性が高いと考えられる

演習問題5-2（10分間）

- 以下の識別法は、生成モデルと識別モデルのどちらに分類されるか？理由も含めて答えよ。
  - ナイーブベイズ識別
  - パーセプトロンの学習規則に基づく識別関数

演習問題5-2（10分間）解答例

- 以下の識別法は、生成モデルと識別モデルのどちらに分類されるか？理由も含めて答えよ。
  - ナイーブベイズ識別
    - 答え：生成モデル
      - $C_{\mathrm{NB}} = \underset{i}{\operatorname{argmax}}\, P(\omega_i) \prod_{j=1}^{d} P\left(x_j | \omega_i\right)$
      - クラスω_iから特徴ベクトル$\boldsymbol{x}$が生成される尤度$P(\boldsymbol{x}|\omega_i)$や事前確率$P(\omega_i)$を用いて事後確率$P(\omega_i|\boldsymbol{x})$を計算
  - パーセプトロンの学習規則に基づく識別関数
    - 答え：識別モデル
      - $g_i(\boldsymbol{x}) = \boldsymbol{w}_i^T \boldsymbol{x}$
      - 識別関数の設計に確率分布などは考慮されていない
      - 与えられた学習データだけで重み$\boldsymbol{w}_i$を学習する

演習問題6-1（10分間）

- 右表のような身長と体重のデータが与えられた
- 身長xから体重$\hat{c}(x)$を予測する線形回帰式が
$$\hat{c}(x) = 0.625x - 48.604$$
であるときの決定係数と相関係数を計算せよ

身長と体重データ

番号	身長 [cm]	体重 [kg]
1	147.9	41.7
2	163.5	60.2
3	159.8	47.0
4	155.1	53.2
5	163.3	48.3
6	158.7	55.2
7	172.0	58.5
8	161.2	49.0
9	153.9	46.7
10	161.6	52.5

演習問題6-1（10分間）解答例

番号	身長 x_i	体重 y_i	$\hat{c}(x)$	$\left(y_i - \hat{c}(x_i)\right)^2$	$(y_i - \tilde{y})^2$
1	147.9	41.7	43.8	4.6	90.8
2	163.5	60.2	53.6	43.8	80.5
3	159.8	47.0	51.3	18.2	17.9
4	155.1	53.2	48.3	23.7	3.9
5	163.3	48.3	53.5	26.6	8.6
6	158.7	55.2	50.6	21.3	15.8
7	172.0	58.5	58.9	0.2	52.9
8	161.2	49.0	52.1	9.9	5.0
9	153.9	46.7	47.6	0.8	20.5
10	161.6	52.5	52.4	0.0	1.6

平均体重 $\tilde{y} \cong 51.2$　　合計：149.0　　合計：297.4

決定係数

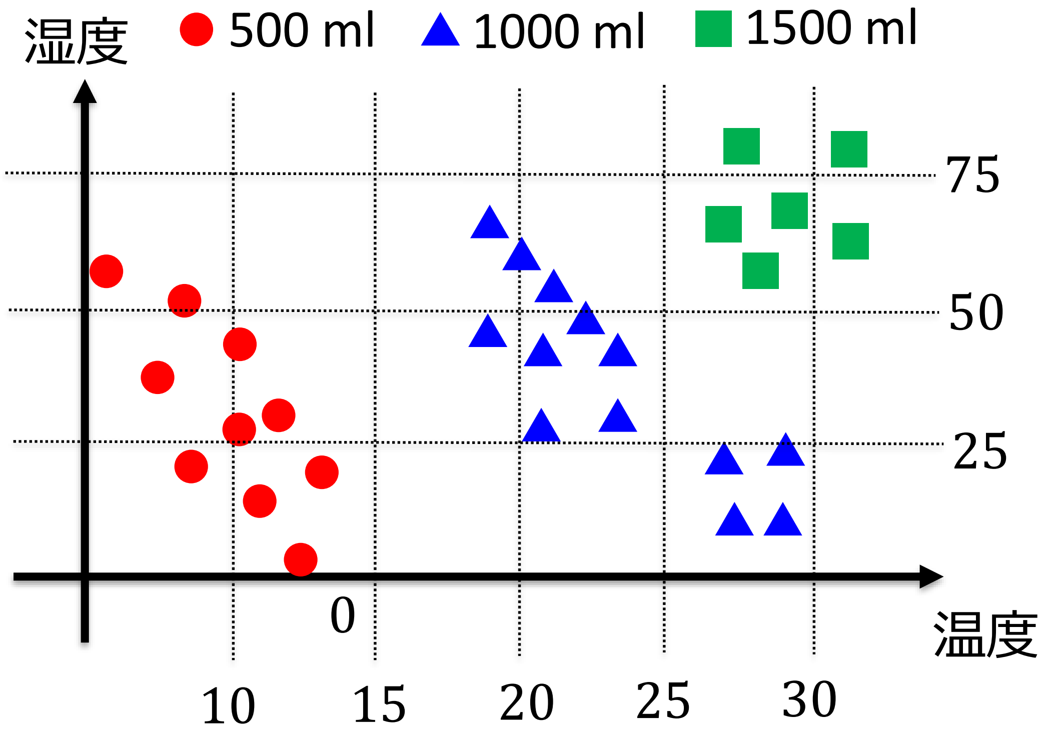
$$R^2 = 1 - \frac{\sum_{i=1}^{N}\{y_i - \hat{c}(x_i)\}^2}{\sum_{i=1}^{N}(y_i - \tilde{y})^2} = 1 - \frac{149.0}{297.4} \cong \mathbf{0.5}$$

相関係数

$$R = \sqrt{R^2} = \sqrt{0.5} \cong \mathbf{0.706}$$

演習問題6-2（10分間）

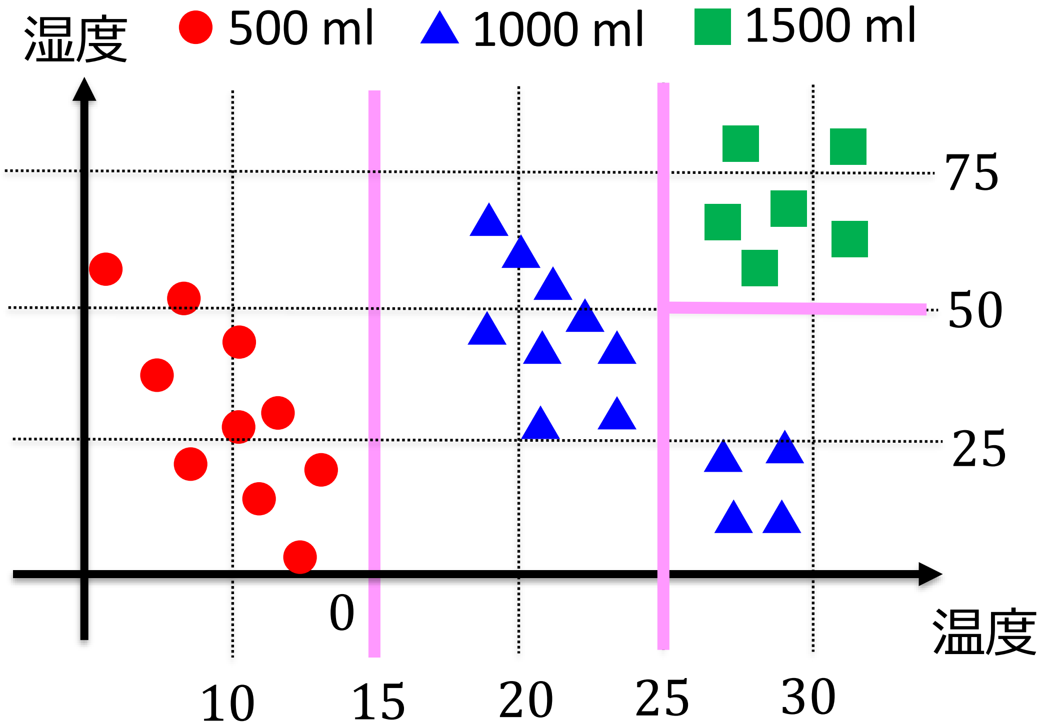
- 以下の学習データが与えられたとき、温度と湿度から1日あたりのビールの消費量を予測する回帰木を作成せよ



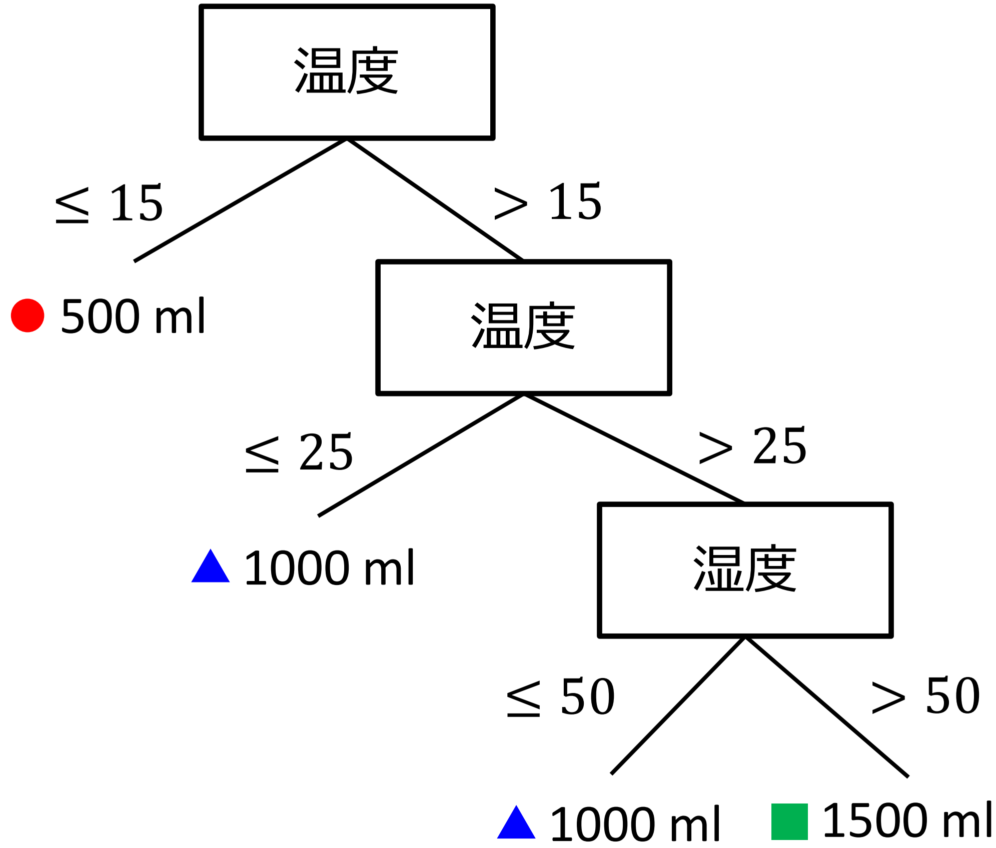
温度・湿度・1日あたりのビールの消費量の関係

演習問題6-2（10分間）解答例

出力値の近いデータが集まるように、
特徴の値によって学習データを分割するのがポイント



温度・湿度・1日あたりのビールの消費量の関係



演習問題7-1（10分間）

- SVMは2クラスのデータに対する分類器である。
ここで、3クラス以上のデータに対する多クラス
識別問題にSVMを適用するには、
どうすれば良いか考えよ

演習問題7-1（10分間）解答例

- SVMは2クラスのデータに対する分類器である。
ここで、3クラス以上のデータに対する多クラス
識別問題にSVMを適用するには、
どうすれば良いか考えよ

- 解答例
  - cクラスの場合（$c > 2$）
あるクラスとそれ以外のデータを識別するc個のSVMを
作成し、結果として最もスコアの高いクラスを選択する

演習問題7-2（10分間）

- カーネル関数の1つである多項式カーネルは、
以下の式で与えられる

$$K(\boldsymbol{x},\boldsymbol{x}')=\boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{\phi}(\boldsymbol{x}')=(\boldsymbol{x}^T\boldsymbol{x}'+1)^p$$

- 2次元特徴ベクトルを $\boldsymbol{x}=(x_1,x_2)$、$p=2$として
$\boldsymbol{\phi}(\boldsymbol{x})$を求めよ

演習問題7-2（10分間）解答例

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + 1)^2$$

$$= (x_1 x_1')^2 + (x_2 x_2')^2 + 2x_1 x_1' x_2 x_2' + 2x_1 x_1' + 2x_2 x_2' + 1$$

$$= \left((x_1)^2, (x_2)^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1\right) \cdot \left((x_1')^2, (x_2')^2, \sqrt{2}x_1' x_2', \sqrt{2}x_1', \sqrt{2}x_2', 1\right)$$

$$\therefore \boldsymbol{\phi}(\boldsymbol{x}) = \left(x_1^2, x_2^2, \sqrt{2}x_1 x_2, \sqrt{2}x_1, \sqrt{2}x_2, 1\right)$$

- 第3項（$\sqrt{2}x_1 x_2$）：特徴の積の項が加わっている
- 2つの特徴が同時に現れるときに大きな値となる

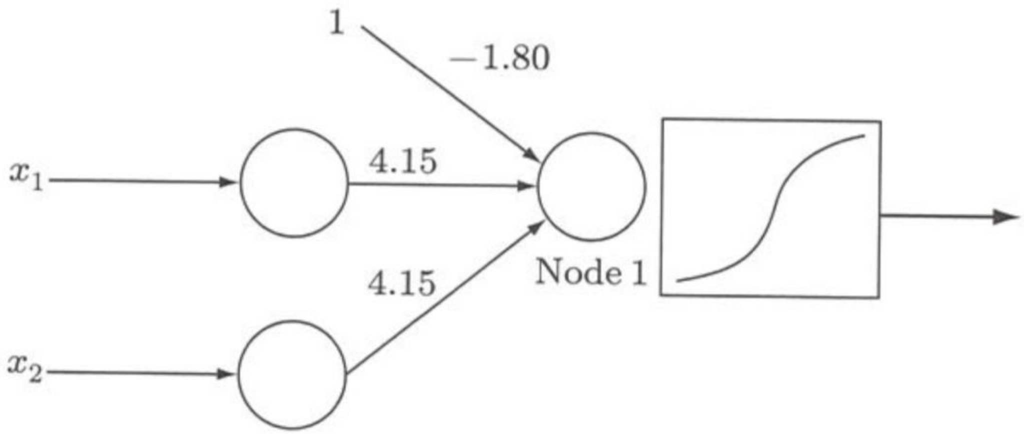
あらわ

→ 共起の情報が加わったということ

きょうき

演習問題8-1（10分間）

- 以下のニューラルネットワークのノードが与えられたとする
  1. 入力x_1とx_2が取りうる値が0と1のどちらかであるとき、x_1とx_2の全ての組み合わせに対する出力を計算せよ
  2. 1.の結果より、どのような入力が与えられると高い出力が得られやすいか考察しなさい



ニューラルネットワークのノード

入力		出力
x_1	x_2	
0	0	
0	1	
1	0	
1	1	

演習問題8-1（10分間）解答例

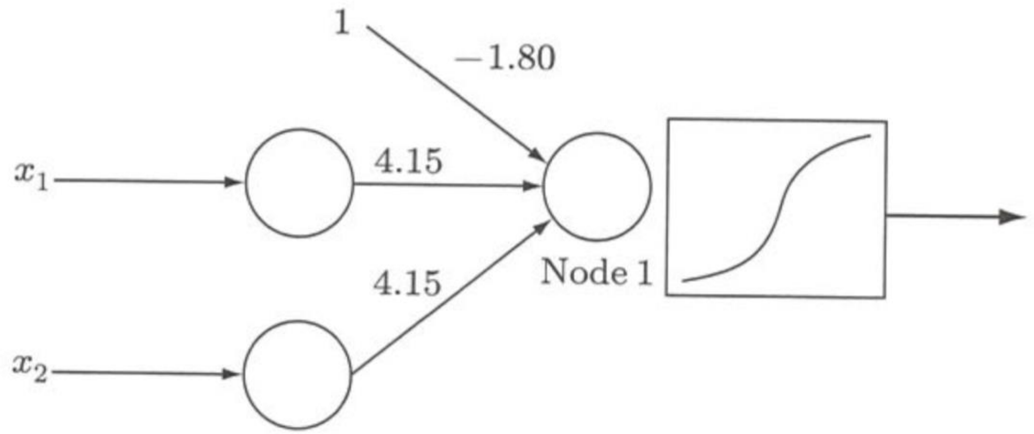
1. 入力x_1とx_2が取りうる値が0と1のどちらかであるとき、x_1とx_2の全ての組み合わせに対する出力を計算せよ

x_1, x_2 に対する出力の値は、以下の式で算出できる

$$f(x_1, x_2) = \frac{1}{1 + \exp\{-(1 \cdot (-1.80) + 4.15 \cdot x_1 + 4.15 \cdot x_2)\}}$$

$$f(x_1 = 0, x_2 = 0) \cong 0.142 \qquad f(x_1 = 0, x_2 = 1) \cong 0.913$$

$$f(x_1 = 1, x_2 = 0) \cong 0.913 \qquad f(x_1 = 1, x_2 = 1) \cong 0.998$$



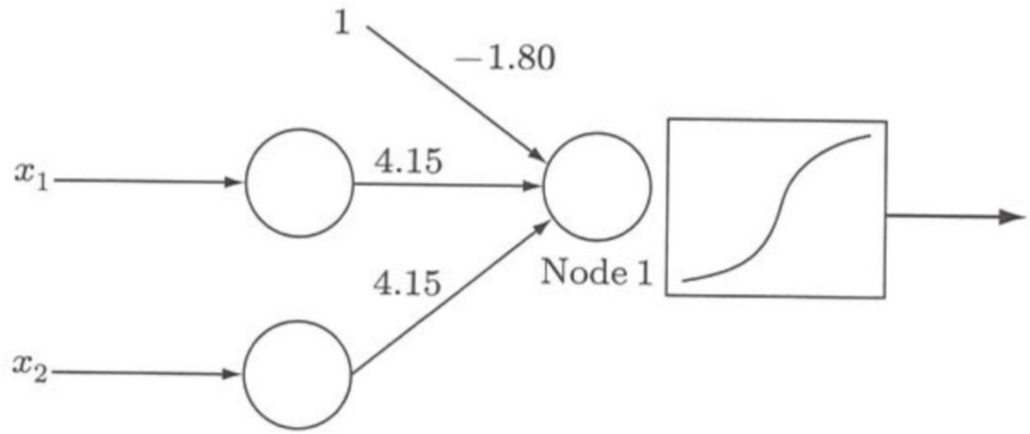
ニューラルネットワークのノード

入力		出力
x_1	x_2	
0	0	0.142
0	1	0.913
1	0	0.913
1	1	0.998

演習問題8-1（10分間）解答例

2. 1.の結果より、どのような入力が与えられると高い出力が得られやすいか考察しなさい

x_1, x_2のどちらかに1が含まれていると高い値が出力される
つまり、このノードは論理演算（OR関数）が表現されている



ニューラルネットワークのノード

入力		出力
x_1	x_2	
0	0	0.142
0	1	0.913
1	0	0.913
1	1	0.998

演習問題8-2（10分間）

- シグモイド関数を活性化関数にすると多層ニューラルネットワークの重みを学習するときに、勾配消失問題が発生(はっせい)しやすい理由を考えよ

演習問題8-2（10分間）解答例

- シグモイド関数を活性化関数にすると
  多層ニューラルネットワークの重みを学習するときに、
  勾配消失問題が発生しやすい理由を考えよ

シグモイド関数の微分値（0に近い値）を使って重みを更新するから

最急勾配法による重みの更新式

$$w_j \leftarrow w_j + \eta \sum_{\boldsymbol{x}_i \in D} (y_i - o_i) o_i (1 - o_i) x_{ij}$$

$$o_i = \frac{1}{1 + \exp(-h_i)}$$

シグモイド関数
（値が取りうる範囲：0〜1）

$$\frac{\partial o_i}{\partial h_i} = \underbrace{o_i}_{1以下}\underbrace{(1 - o_i)}_{1以下}$$

シグモイド関数の微分
1以下の値の積（0に近づく）

演習問題9-1（10分間）

- 以下の0から7までの数を自己写像する
ニューラルネットワーク（オートエンコーダ）を考える
  - 入力層・出力層：8次元ベクトル
    - 表現する数に対応する特徴のみが1、その他は0
      - 例：「3」という数に対する入力ベクトルは「0,0,1,0,0,0,0,0」
  - 中間層のノード数：3

- このネットワークを学習した結果、
中間層では、入力がどのように表現されているか
（入力からどのような情報が獲得されたか）考えよ

演習問題9-1（10分間）解答例

- このネットワークを学習した結果、中間層では、
  入力がどのように表現されているか考えよ

- 解答例
  - 「**2進(しん)エンコーディングの概念**」が獲得されている（かも）
  - 0から7の数は、3桁以内(けたいない)の2進数で表現できる
    - 8ノードの入力情報が、より少ない3ノードの中間情報で表現
    - 「3」を例に考えると
      - 入力層：「0,0,1,0,0,0,0,0」、中間層（2進数）：「0,1,1」

オートエンコーダは、このようにして得られた中間層の値を入力として、1階層上にずらして、同様の表現学習を行う。この手順を積み重ねると、入力に近い側では単純な特徴が、階層が上がっていくにつれて複雑な特徴が学習される

演習問題9-2（10分間）

- 畳み込みニューラルネットワークを用いたモデルを構築することを考える。このネットワークを構築するために、事前に決めておくべきパラメータは何か考えよ
  - 1つ目は「活性化関数を何にするか？」
  - それ以外に事前に決めておかないといけないものは何？

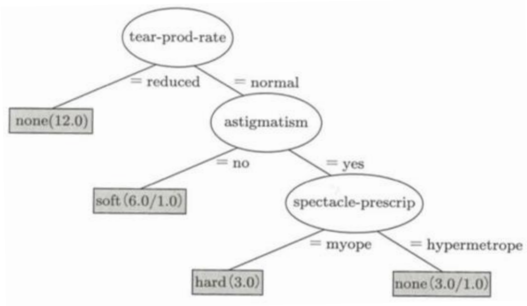
演習問題9-2（10分間）解答例

- 畳み込みニューラルネットワークを用いたモデルを構築することを考える。このネットワークを構築するために、事前に決めておくべきパラメータは何か考えよ
  - 活性化関数
  - 畳み込み層のフィルタサイズとフィルタ数
  - プーリングの種類（平均なのか？最大値なのか？）
  - 層の構成（畳み込み層、全結合層の数など）
  - ドロップアウトの割合
  - 重みを更新する式の学習係数（最適化器）
  - 更新回数
  - 事前学習の有無　　など

演習問題10-1（10分間）

- バギングで作成する識別器は不安定な方が良い。このアルゴリズムで用いる不安定な識別器は、どのようなものがあるか考えよ。

演習問題10-1（10分間）解答例

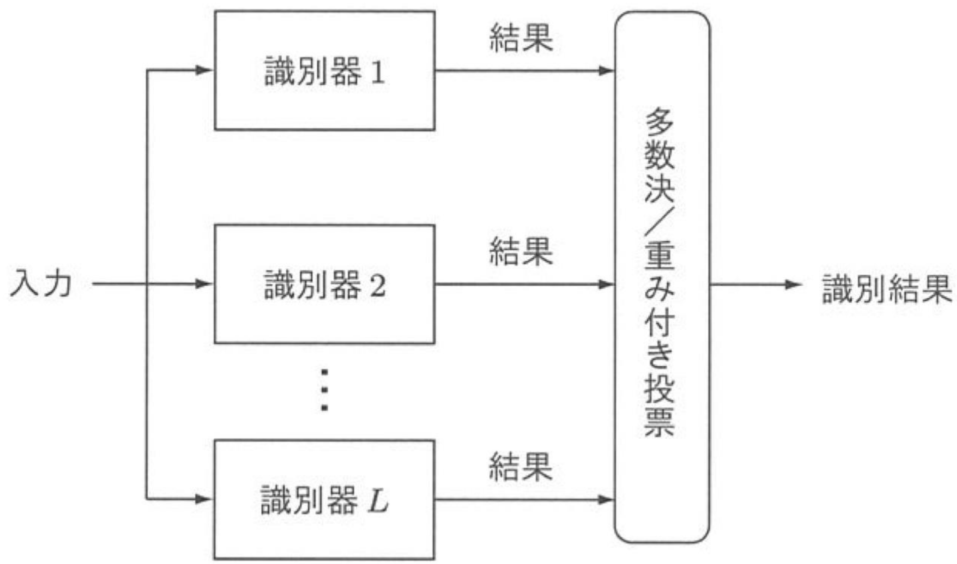


■ 解答例

- 決定木
  - 得られる情報が多い特徴を木の上方のノードに配置するように構成
- 決定木は過学習を起こしやすい
  - 過学習は、モデルが学習データに特化しすぎて、未知データに対して性能が下がる現象（げんしょう）
- 全ての事例のエラーが無くなるように決定木を作成すると、木が成長しすぎて、学習データに適応しすぎた過学習になりがち
  - 言い換えると、学習データを変えると、決定木の構成も異なる

演習問題10-2（10分間）

- この講義ではアンサンブル学習によって識別問題を解く方法を紹介した。これを回帰問題に適用するにはどうすれば良いか考えなさい。



アンサンブル学習による識別問題

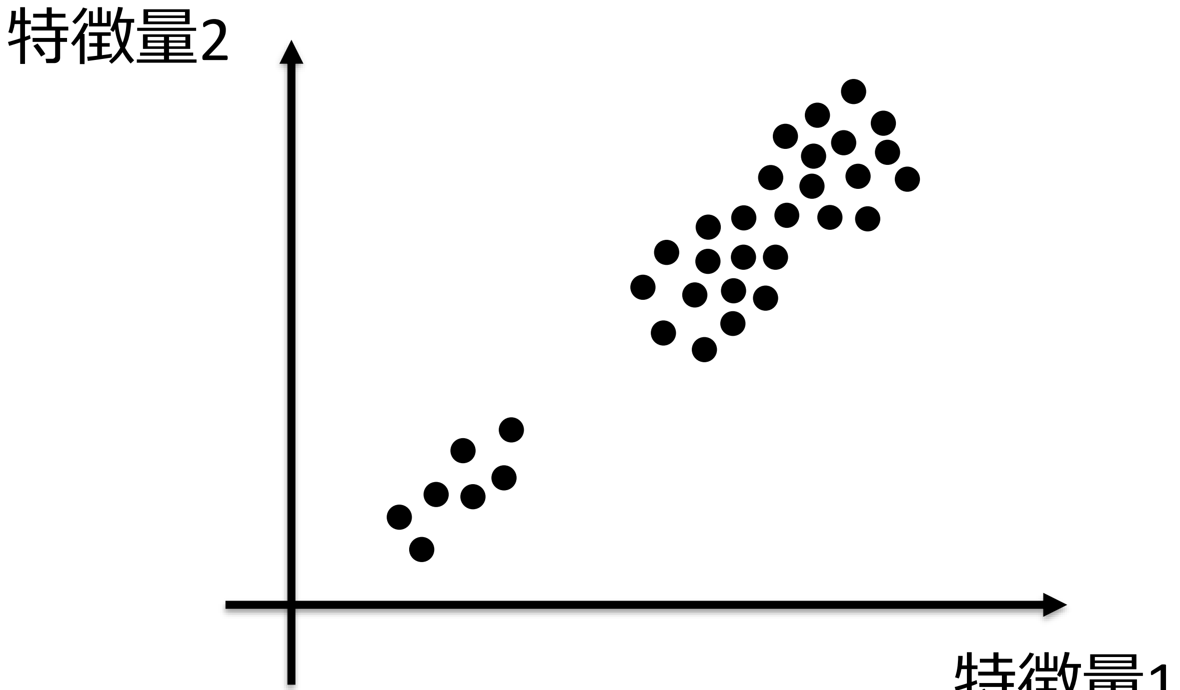
演習問題10-2（10分間）解答例

- この講義ではアンサンブル学習によって識別問題を解く方法を紹介した。これを回帰問題に適用するにはどうすれば良いか考えなさい。

- 解答例
  - アンサンブル学習で、複数の回帰モデルを生成
  - それぞれのモデルの出力値の平均値や中央(ちゅうおう)値を用いて最終的な回帰の結果を得る

演習問題11-1（10分間）

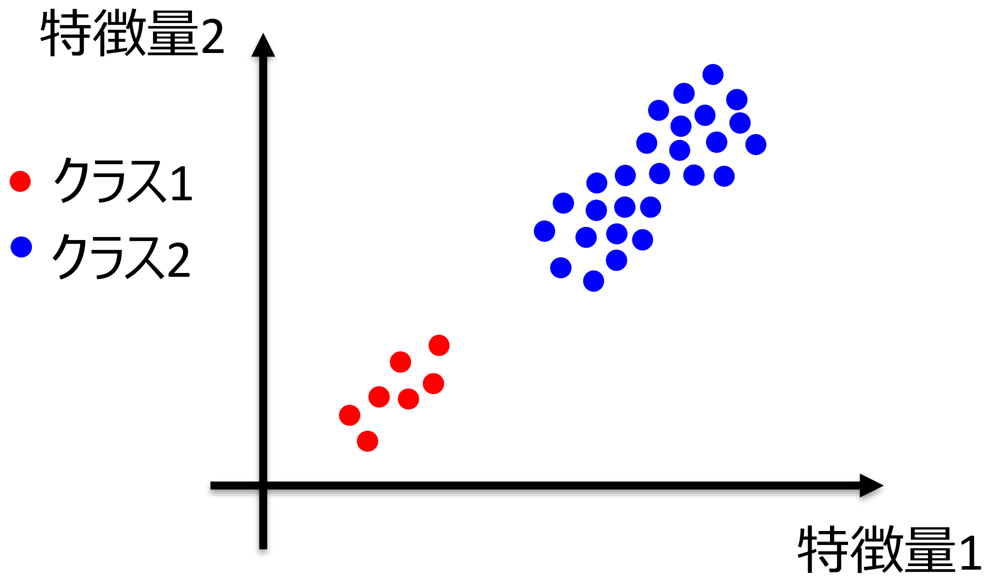
- 以下の教師なしデータに対して、分割数が2の場合と3の場合のそれぞれで分割最適化クラスタリングを適用したとき、どのような結果が期待されるか？

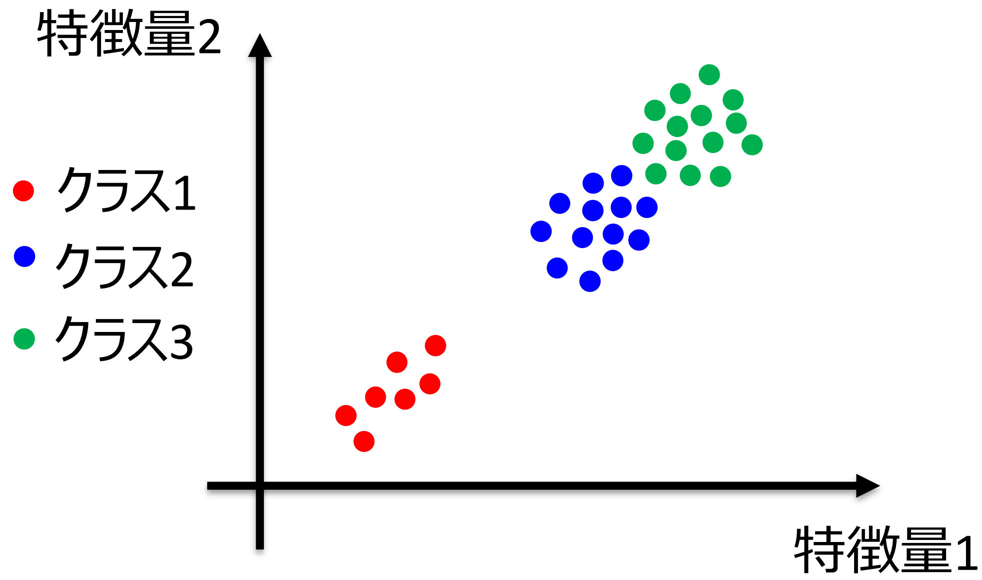


演習問題11-1（10分間）解答例

- 以下の教師なしデータに対して、分割数が2の場合と3の場合のそれぞれで分割最適化クラスタリングを適用したとき、どのような結果が期待されるか？

- 解答例





演習問題11-2（10分間）

- k-meansアルゴリズムのクラスタリング結果を用いて新たなデータが観測されたときに、そのデータが属するクラスタを決める方法を考えなさい

- k-meansアルゴリズムのクラスタリング結果に基づく教師なし学習では、良い識別器を作ることができない理由を考えなさい

演習問題11-2（10分間）解答例

- k-meansアルゴリズムのクラスタリング結果を用いて新たなデータが観測されたときに、そのデータが属するクラスタを決める方法を考えなさい

- 解答例
  - k-meansアルゴリズムの結果は、各クラスの平均ベクトルなので、新しく観測したデータと、各クラスタの平均ベクトルとの距離を求めて、最も近いクラスタを分類結果とする

演習問題11-2（10分間）解答例

- k-meansアルゴリズムのクラスタリング結果に基づく教師なし学習では、良い識別器を作ることができない理由を考えなさい

- 解答例
  - この距離計算は、全クラスタの分散が等しいことを前提にしており、クラスタごとにデータの広がり方が異なる場合には適切な結果とならない
  - クラスタの事前確率も考慮されていない
  - 事後確率を計算する確率モデルを作り、その確率が最大となるモデルパラメータをEMアルゴリズムで学習する

演習問題12-1（10分間）

- 以下のスーパーマーケットの売り上げ記録がある
  - 商品点数：{ミルク、パン、バター、雑誌（ざっし）} の4点
  - トランザクション：6件（6回分の購入記録）

1. このデータは「疎らなデータ」、「密なデータ」のどちらであるか考えよ
2. 項目集合{ミルク、パン}の支持度を示すsupport({ミルク、パン})を求めなさい
3. 商品点数が1000種類であったとき、可能な項目集合の数はいくつか？
   ヒント：商品は「買った」「買っていない」の2種類に分類できると考える

No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

スーパーマーケットの売り上げ記録
（tはその商品が買われたことを意味する）

演習問題12-1（10分間）解答例

1. このデータは「疎らなデータ」、「密なデータ」のどちらであるか考えよ

□ 解答例

- 疎らなデータ
- トランザクションの特徴ベクトルのほとんどの次元の値は空白で、少数の次元のデータが埋まっている
  - スーパーマーケットで揃えている商品の種類数：数千～数万
  - 1回の買い物で1人の客が買う商品の種類数：多くて数十

スーパーマーケットの売り上げ記録

No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

演習問題12-1（10分間）解答例

2. 項目集合{ミルク、パン}の支持度を示す support({ミルク、パン})を求めなさい

解答例

- support({ミルク、パン}) $= \frac{T_{\{\mathrm{items}\}}}{T} = \frac{3}{6} = \mathbf{0.5}$
  - T ：全トランザクション件数＝6
  - T_{items} ：項目集合itemsが出現するトランザクション件数
    - $T_{\{\mathrm{items}\}} = T_{\{ミルク、パン\}} = 3$

No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

スーパーマーケットの売り上げ記録

演習問題12-1（10分間）解答例

3. 商品点数が1000種類であったとき、
可能な項目集合の数はいくつか？

□ 解答例

- ■ **$2^{1000} - 1$通り**
  - 1000点の商品に対して{買った、買わなかった}の値を割(わ)り振(ふ)るので、
2^{1000}通りとなり、この中から何も買われなかった場合の1通りを除く
- ■ この結果からも、まともに全ての可能な組み合わせについて計算することは難しいことがわかる
  - 支持度を計算する項目集合をいかにして絞り込むかが重要！

演習問題12-2（5分間）

- 「ハム→卵(たまご)」という規則について、以下の情報が得られたとする
  - 項目集合 {ハム、卵} の支持度が0.1
  - 「ハム→卵」という規則の確信度が0.7、リフト値：5

- 上記の条件において、以下の項目を求めなさい
  1. ハムと卵を同時に購入している客の割合
  2. ハムを購入した客の中で卵も購入する客の割合
  3. 「ハムを既(すで)に買った客が卵を買う確率」は「任意の客が卵を買う確率」の何倍か？

演習問題12-2（5分間）解答例

- 「ハム→卵」という規則について、以下の情報が得られたとする
  - 項目集合 {ハム、卵} の支持度が0.1
  - 「ハム→卵」という規則の確信度が0.7、リフト値：5

- 解答例
  1. ハムと卵を同時に購入している客の割合：10 %（支持度）
  2. ハムを購入した客の中で卵も購入する客の割合：70 %（確信度）
  3. 「ハムを既に買った客が卵を買う確率」は「任意の客が卵を買う確率」の何倍か？：5倍（リフト値）

演習問題12-3（10分間）

- 実際に協調フィルタリングが用いられている事例を調べなさい

演習問題12-3（10分間）解答例

- 実際に協調フィルタリングが用いられている事例を調べなさい

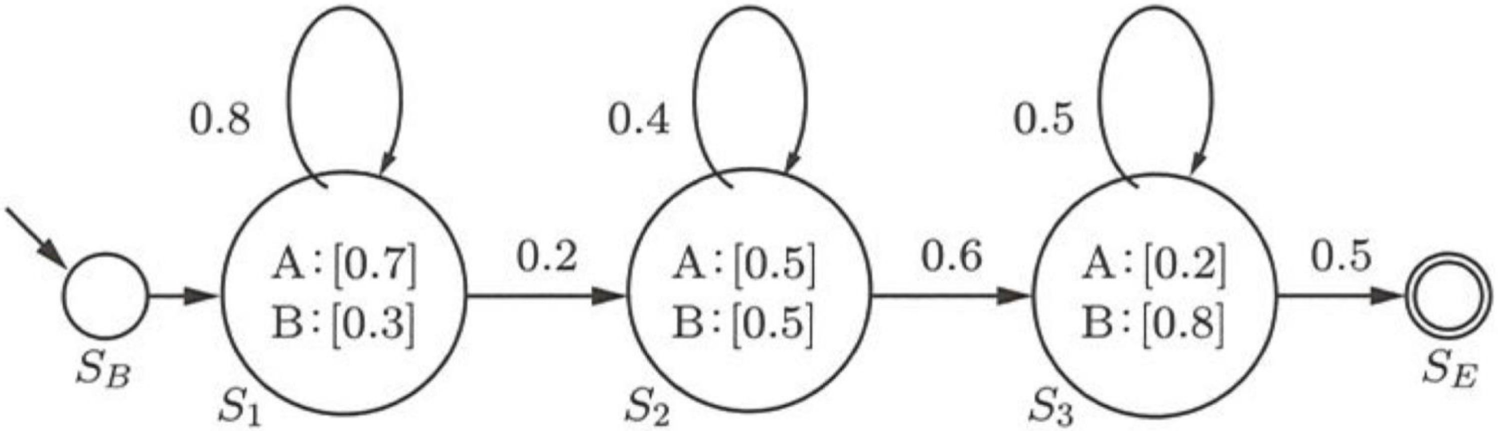
- 解答例
  - Amazonオンラインショップ
  - ある商品を検索（けんさく）すると
    - 同時に購入される商品
    - 他にチェックしている商品

  が一緒に表示（ひょうじ）される

ページの下の方に
進むと…

演習問題13-1（10分間）

- 下図に示すHMMが与えられているとき、特徴ベクトル系列 "AAABB" が出力されるような状態遷移系列を全て求めよ



HMM

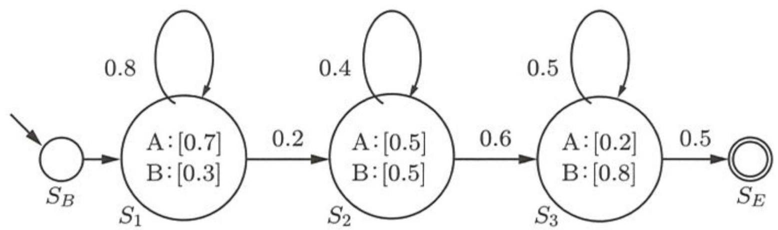
【ヒント】

- 1つは $S_B \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(B)}{S_2} \rightarrow \overset{(B)}{S_3} \rightarrow S_E$
- 他にどのような状態遷移系列が考えられるか？

演習問題13-1（10分間）解答例

- 下図に示すHMMが与えられているとき、
  特徴ベクトル系列 “AAABB” が出力されるような
  状態遷移系列を全て求めよ

- 解答例
  - 系列 “AAABB” が出力されるということは、以下の制約がつく
    - 最初の記号Aは状態S_1から出力
    - 最後の記号Bは状態S_3から出力
  - このような制約の中で、
    考えられる状態遷移を全て洗い出せばよい

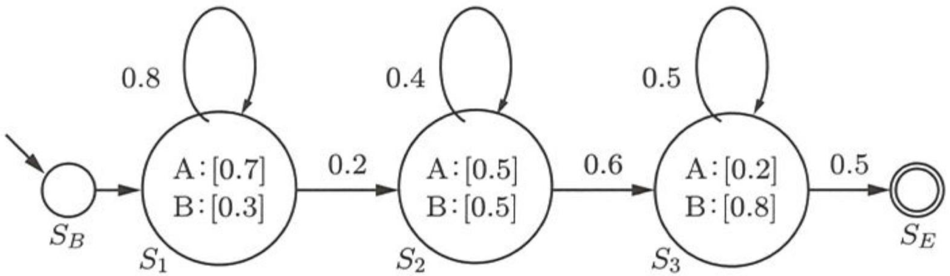


演習問題13-1（10分間）解答例

□ 解答例（続き）

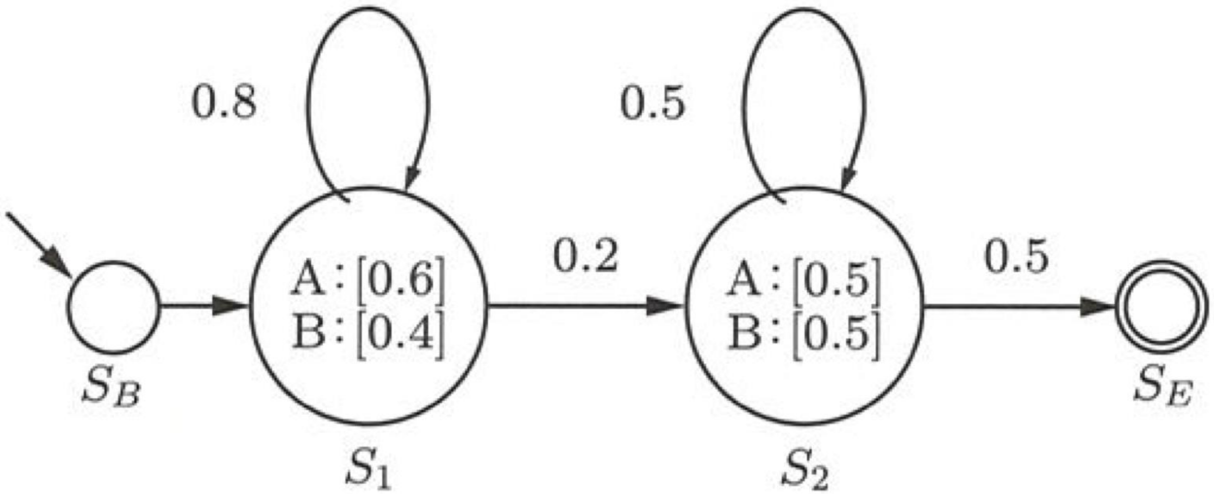
- 系列 “AAABB” を出力できる状態遷移系列は以下の6通り
  - $S_B \rightarrow S_1 \rightarrow S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_E$
  - $S_B \rightarrow S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_2 \rightarrow S_3 \rightarrow S_E$
  - $S_B \rightarrow S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_3 \rightarrow S_E$
  - $S_B \rightarrow S_1 \rightarrow S_2 \rightarrow S_2 \rightarrow S_2 \rightarrow S_3 \rightarrow S_E$
  - $S_B \rightarrow S_1 \rightarrow S_2 \rightarrow S_2 \rightarrow S_3 \rightarrow S_3 \rightarrow S_E$
  - $S_B \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_3 \rightarrow S_3 \rightarrow S_E$

    (A) (A) (A) (B) (B)



演習問題13-2（15分間）

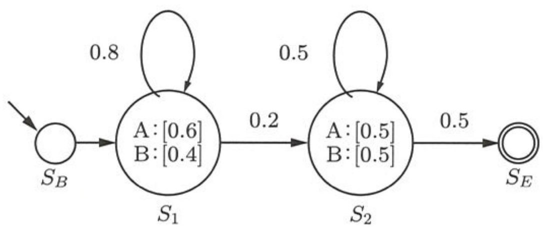
- 下図に示すHMMが与えられているとき、
  特徴ベクトル系列 “AAB” の出力確率を求めよ



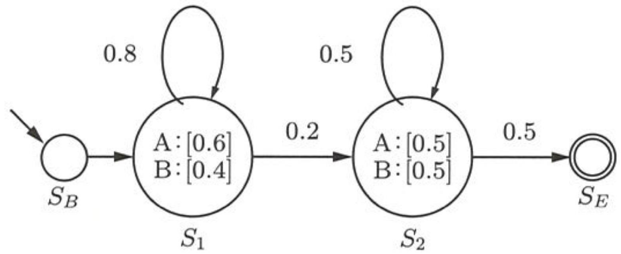
HMM

演習問題13-2（15分間）解答例

- 下図に示すHMMが与えられているとき、特徴ベクトル系列 “AAB” の出力確率を求めよ
- 解答例
  - 系列 “AAB” が出力されるということは、以下の制約がつく
    - 最初の記号Aは状態S_1から出力
    - 最後の記号Bは状態S_2から出力
  - 系列長: 3、状態数: 2より、S_1とS_2のどちらかで1回自己遷移
  - 系列 “AAB” を出力できる状態遷移系列は以下の2通り
    - $S_B \rightarrow S_1 \rightarrow S_1 \rightarrow S_2 \rightarrow S_E$
    - $S_B \rightarrow \underset{(A)}{S_1} \rightarrow \underset{(A)}{S_2} \rightarrow \underset{(B)}{S_2} \rightarrow S_E$



演習問題13-2（15分間）解答例



□ 解答例（つづき）

- $S_B \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(B)}{S_2} \rightarrow S_E$ の確率

$$P(S_B \rightarrow S_1)P(A|S_1)P(S_1 \rightarrow S_1)P(A|S_1)P(S_1 \rightarrow S_2)P(B|S_2)P(S_2 \rightarrow S_E)$$
$$= 1.0 \times 0.6 \times 0.8 \times 0.6 \times 0.2 \times 0.5 \times 0.5 = 0.0144$$

- $S_B \rightarrow \overset{(A)}{S_1} \rightarrow \overset{(A)}{S_2} \rightarrow \overset{(B)}{S_2} \rightarrow S_E$ の確率

$$P(S_B \rightarrow S_1)P(A|S_1)P(S_1 \rightarrow S_2)P(A|S_2)P(S_2 \rightarrow S_2)P(B|S_2)P(S_2 \rightarrow S_E)$$
$$= 1.0 \times 0.6 \times 0.2 \times 0.5 \times 0.5 \times 0.5 \times 0.5 = 0.0075$$

- これらは同時には起こりえない事象なので、
  最終的に求める確率は、これらの和となる
- 系列 “AAB” の出力確率 $= 0.0144 + 0.0075$ $= \mathbf{0.0219}$

演習問題14-1（5分間）

- 「強化学習」と「教師あり/教師なし学習」の違いを考えなさい

演習問題14-1（5分間）解答例

- 「強化学習」と「教師あり/教師なし学習」の違いを考えなさい
- 解答例
  - 教師信号が間接的
    - 「何が正解か？」ではなく、報酬（連続値）が時々与えられる
  - 報酬が遅(おく)れて与えられる
    - 将棋の勝利、迷路のゴール
  - 探求(たんきゅう)が可能
    - エージェントが自分で学習の対象を選べる
  - 状態が確定(かくてい)的でない場合がある
    - 確率分布でそれぞれの状態にいる確率を表せる

演習問題15-1（10分間）

- 人間の成長過程は、半教師あり学習に似ていると言われている。その理由を、実際の事例を挙(あ)げながら述べなさい。

演習問題15-1（10分間）解答例

- 人間の成長過程は、半教師あり学習に似ていると言われている。その理由を、実際の事例を挙げながら述べなさい。

- 解答例
  - 幼(おさな)い頃(ころ)、大人(おとな)たちに「これは猫(ねこ)だよ」、「これは犬(いぬ)だよ」と何度か教えてもらっただけで、あとは自分で大量の動物を見ながら、猫や犬を正確に認識していく
    - 大人たちから教えてもらった少量の情報 → 正解付きデータ
    - 自分自身が見た大量の動物 → 正解なしデータ

演習問題15-2（10分間）

- 半教師あり学習を活用できる場面を考えなさい

演習問題15-2（10分間）解答例

- 半教師あり学習を活用できる場面を考えなさい

- 解答例
  - 以下のような条件を満たす状況を考えるのがポイント
    - 用意できる正解データが少ない
    - 収集したデータに正解情報を付与するのに大きなコストがかかる
  - 医療用画像の分類・認識
    - ラベル付けされた医療用画像は少量
      - その病気と診断された患者の数だけしか集められない
    - 画像に正解情報を付与できる人が限られる（コストが大きい）
      - 基本的に医療用画像から病気を診断できるのは医師（専門家）だけ