目次

# 最適化と認識・学習

## 文殊の知恵

高橋那弥

## 1　一般的な制約付き最適化問題

この問題の定義は以下のようなものである。$X$ を $R^n$ の線形空間とし、$A$ はその部分空間とする。
また、$f(x), g_k(x), k=1,...,m$ は $R^n \to R$ の関数とする。ここで

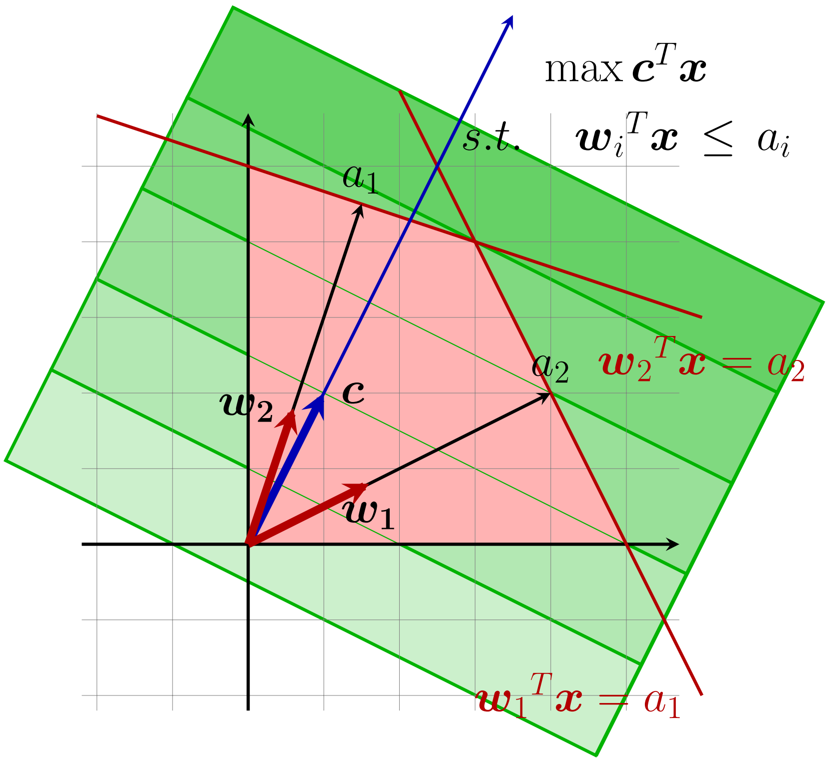
$$\begin{aligned}
&\quad \boldsymbol{x} \in A \ \text{:設計変数} \\
&\quad g_k(x) \le 0,\ \ k = 1, 2, \ldots, m \ \text{:制約条件} \\
&\text{の下で} \\
&\quad f(x) \ \ \text{:目的関数} \\
&\quad \text{を最小化（あるいは最大化）する問題}
\end{aligned}$$

### 1.1　線形計画問題

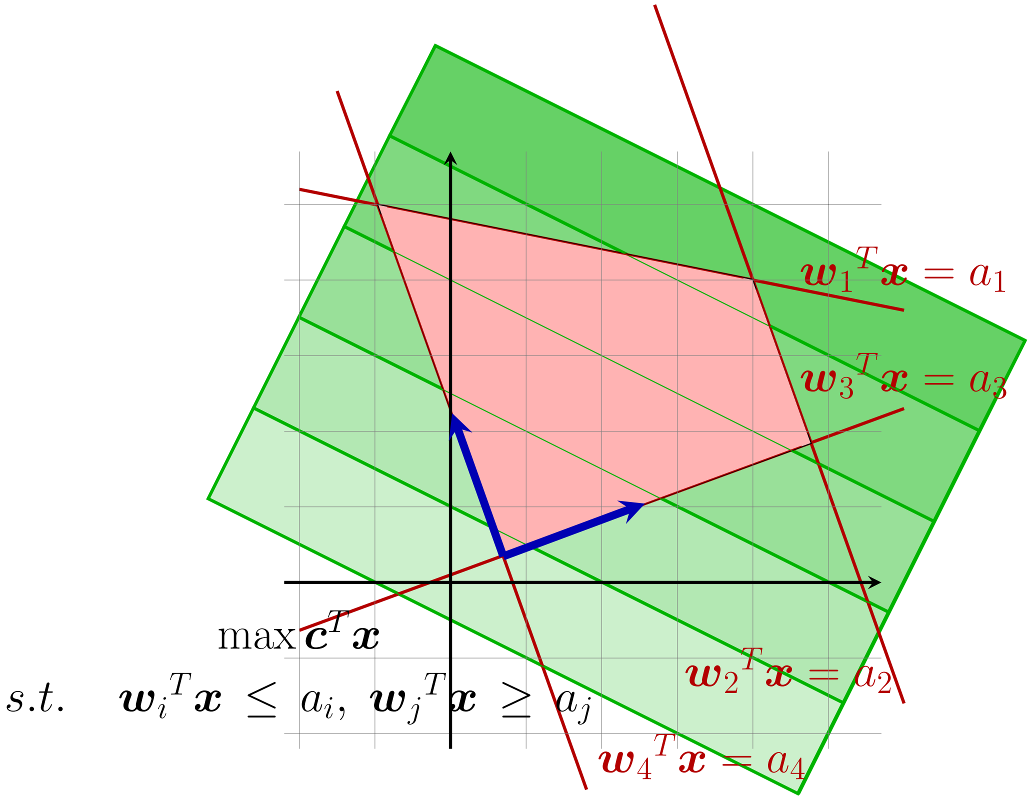
制約付きの最適化問題で $f(x), g(x)$ がともに以下のように線形関数の時、線形計画問題と呼ばれる。

$$\begin{aligned}
&\quad \boldsymbol{x} = (x_i),\ x_i \ge 0 \ \text{:設計変数} \\
&\quad g_k(x) = {\boldsymbol{w}_k}^T \boldsymbol{x} \le \alpha_k,\ \alpha_k \ge 0,\ \ k = 1, 2, \ldots, m \ \text{:制約条件} \\
&\text{の下で} \\
&\quad f(x) = \boldsymbol{c}^T \boldsymbol{x} \ \ \text{:目的関数} \\
&\quad \text{を最小化（あるいは最大化）する問題}
\end{aligned}$$

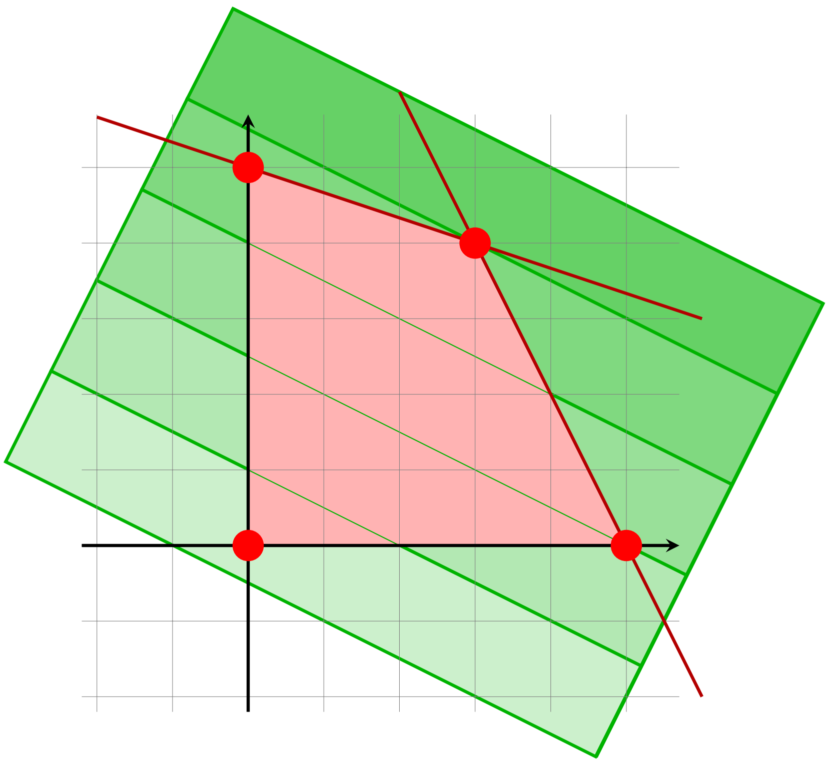
以下のような問題のことを表す。



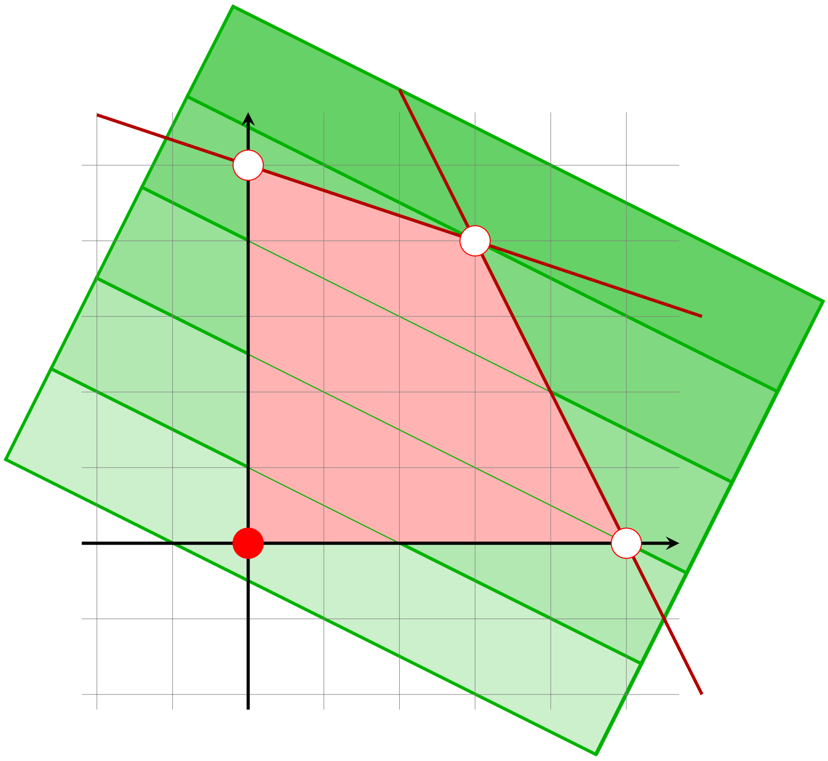
以下のような一般的な線形問題は適当な座標変換により、前図のような形式に変換できるので、ここでは、前図の場合を考える。



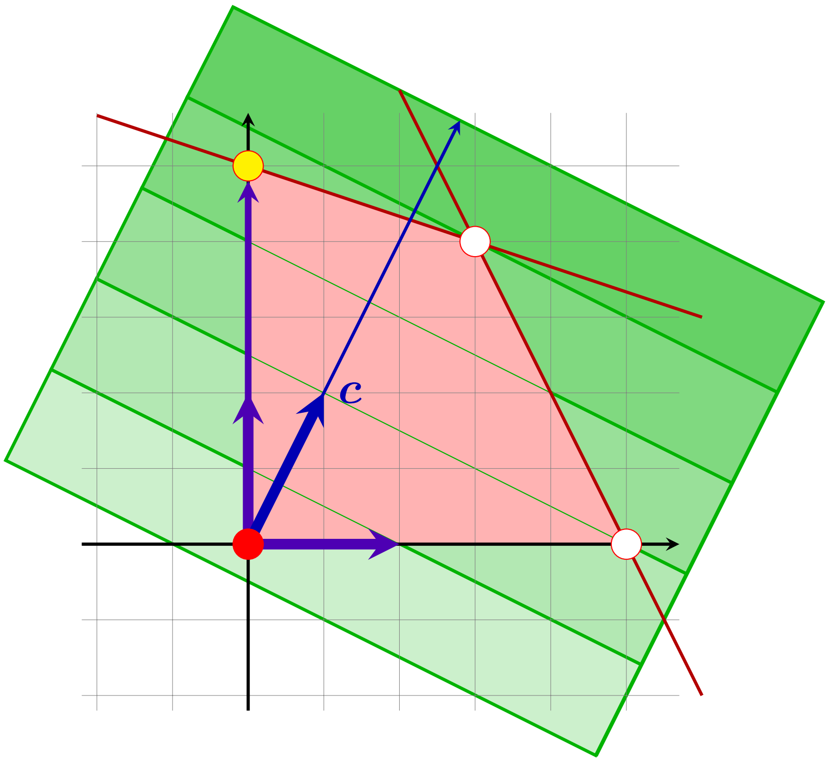

よって、線形計画問題では必ず制約条件が作る多角形の頂点にいずれかにおいて最適解をもつ。



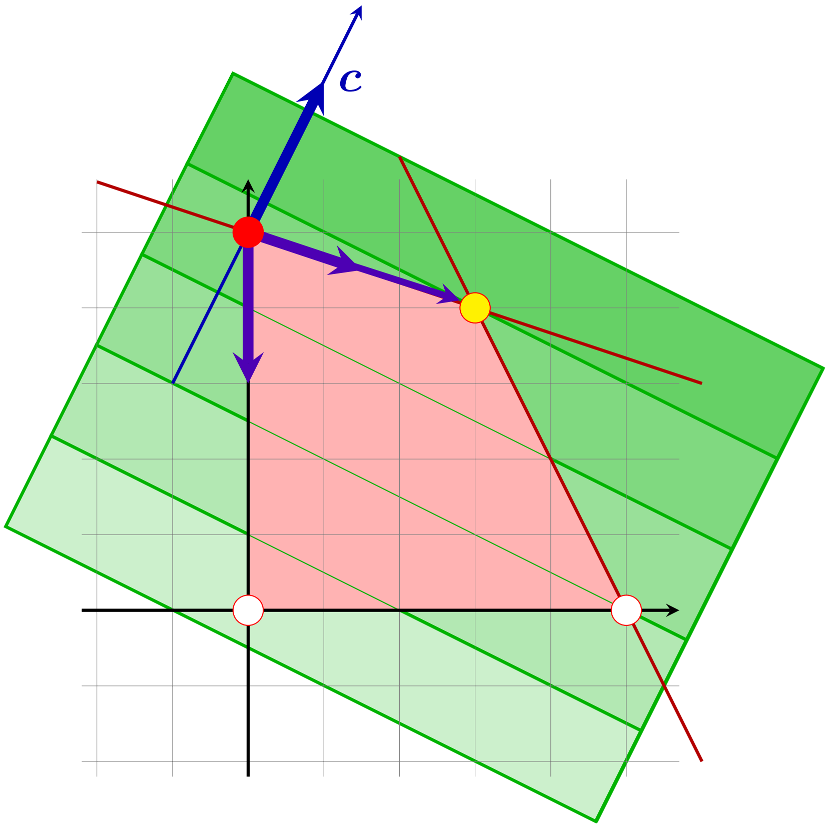
原点は可能解であるので、原点を初期解として、これを改善することを考える。



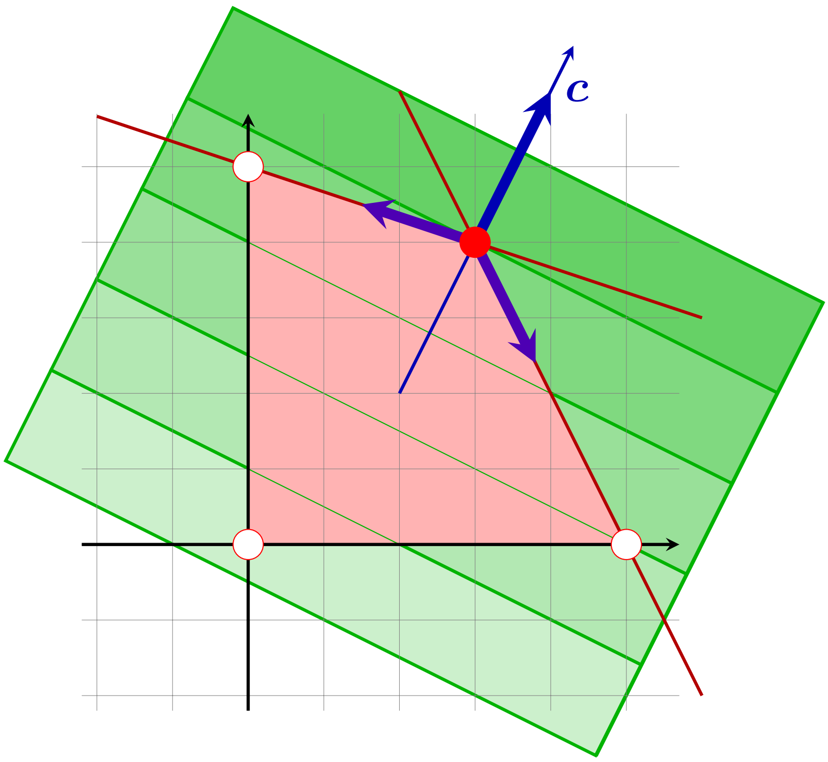
暫定解から伸びる辺の方向に解を改善することを考える。暫定解から複数ある候補に向かう時急な正の勾配を持つ方向を選んで解を改善する。



この時、勾配はどちらに行っても負となるため、改善する方向が見つからなくなり。最適解と考えられるため、終了する。




処理を繰り返す。




この時、勾配はどちらに行っても負となるため、改善する方向が見つからなくなり。最適解と考えられるため、終了する。

### 1.2　シンプレックス法

以上のような線形計画問題を解く解法の一つとして、シンプレックス法が存在し、以下のような解法である。

1. 正準形に変換する。
   正準形とは制約条件式が等式となっているもののことをいう。
2. 以下を繰り返す。
   基底変数とは独立変数のことであり、制約条件式の中で一つしか出てこないもののことである。
   - 2-1 全ての基底変数を 0 とおいたときの目的関数の値（基底可能解）を $z$ とする。
   - 2-2 基底変数の内最適化に最も寄与するものを選び、その基底変数を $x$ とする。寄与するものがなければ現在の $z$ を最適値として終了する。
   - 2-3 $x$ 以外の基底変数を 0 のままとし、$x$ だけを変化させるとき、どの条件式まで $x$ を変化させることができるかを調べる。この時選ばれる条件式を $S$ とする。
   - 2-4 $S$ を用いて、他の条件式から $x$ を消去して、基底変数を入れ替える。（$x$ を基底変数から外す。この時非基底変数のどれかが基底変数となる。）

### 1.3　具体例

以下の問題を考える。

$$\begin{array}{rcrcl} x_1 & + & 3x_2 & \leq & 15 \\ 2x_1 & + & x_2 & \leq & 10 \\ x_1 & + & 2x_2 & = & z \end{array} \tag{1.1}$$

の時、条件 $x_1 \geq 0, x_2 \geq 0$ のもとで式 (1.1) を満たすような $z$ を最大化する問題を考える。この時、まず不等式を等式に変えるために $y_1 \geq 0, y_2 \geq 0$ のスラック変数 $y_1, y_2$ を導入して以下のように式変形する。この時、$\boldsymbol{w}_1 = [\,1\ 3\,]^T$, $\boldsymbol{w}_2 = [\,2\ 1\,]^T$, $\boldsymbol{c} = [\,1\ 2\,]^T$ と表される。

$$\begin{array}{rcrcrcrcrcl} x_1 & + & 3x_2 & + & y_1 & & & & & = & 15 \\ 2x_1 & + & x_2 & & & + & y_2 & & & = & 10 \\ x_1 & + & 2x_2 & & & & & - & z & = & 0 \end{array} \tag{1.2}$$

この変形により、式 (1.1) は条件 $x_1 \geq 0, x_2 \geq 0, y_1 \geq 0, y_2 \geq 0$ のもとで式 (1.2) を満たすような $z$ を最大化する問題となる。

この時、式 (1.2) は正準形であるため、シンプレックス法を用いて解答することができる。

よって、以下のように回答できる。

まず、式 (1.2) においては独立変数は $x_1, x_2$ であるため、これらを全て 0 にすると以下のようになる。

$$\begin{array}{lcrcrcrcrcl} 0 & + & 3\cdot 0 & + & y_1 & & & & & = & 15 \\ 2\cdot 0 & + & 0 & & & + & y_2 & & & = & 10 \\ 0 & + & 2\cdot 0 & & & & & - & z & = & 0 \end{array}$$

この時、制約条件を満たすので、$(x_1, x_2) = (0, 0)$ は基底可能解である。この時、$z = 0$

次に正の勾配で大きいほうを選んで更新するので、長さが 1 で方向が $x_2$ 軸方向のベクトルを $\boldsymbol{t}_y$ とし、$x_1$ 軸方向のベクトルを $\boldsymbol{t}_x$ とおくと、これらと $\boldsymbol{c}$ の内積はそれぞれ式 (1.2) の $x_1, x_2$ の係数となるので、係数を比べて大きいほうに解を改善するので、$x_2$ 軸方向に更新する。

よって、基底変数である $x_1$ を 0 とし、$x_2$ をどこまで動かせるか調べる。

$$
\begin{array}{lllllllllll}
0 & + & 3x_2 & + & y_1 & & & & & = & 15 \\
2\cdot 0 & + & x_2 & & & + & y_2 & & & = & 10 \\
x_1 & + & 2\cdot x_2 & & & & & - & z & = & 0
\end{array}
$$

$$
\Longrightarrow
\begin{array}{lllllllllll}
0 & + & 3x_2 & + & 0 & & & & & \leq & 15 \\
2\cdot 0 & + & x_2 & & & + & 0 & & & \leq & 10 \\
x_1 & + & 2\cdot x_2 & & & & & - & z & = & 0
\end{array}
$$

$$
\Longrightarrow
\begin{array}{lllllllllll}
\frac{1}{3}\cdot 0 & + & x_2 & + & \frac{1}{3}\cdot 0 & & & & & \leq & 5 \\
2\cdot 0 & + & x_2 & & & + & 0 & & & \leq & 10 \\
x_1 & + & 2\cdot x_2 & & & & & - & z & = & 0
\end{array}
$$

よって、この時、$x_2$ は 5 まで動かせることを意味し、$\frac{1}{3}\cdot 0 + x_2 + \frac{1}{3}\cdot 0 \leq 5$ が条件式としてえらばれる。よって、以下のように変形できる。

$$
\begin{array}{lllllllllll}
\frac{1}{3}x_1 & + & x_2 & + & \frac{1}{3}y_1 & & & & & = & 5 \\
2x_1 & + & x_2 & & & + & y_2 & & & = & 10 \\
x_1 & + & 2x_2 & & & & & - & z & = & 0
\end{array}
$$

よって、条件式 $\frac{1}{3}x_1 + x_2 + \frac{1}{3}y_1 = 5$ を用いて、$x_2$ を基底変数から外すと以下のように変形できる。

$$
\begin{array}{lllllllllll}
\frac{1}{3}x_1 & + & x_2 & + & \frac{1}{3}y_1 & & & & & = & 5 \\
\frac{5}{3}x_1 & & & - & \frac{1}{3}y_1 & + & y_2 & & & = & 5 \\
\frac{1}{3}x_1 & & & - & \frac{2}{3}y_1 & & & - & z & = & -10
\end{array}
$$

この時、$x_2$ を基準にして、基底変数が $x_1, y_1$ となる。

$$
\begin{array}{lllllllllll}
\frac{1}{3}\cdot 0 & + & x_2 & + & \frac{1}{3}\cdot 0 & & & & & = & 5 \\
\frac{5}{3}\cdot 0 & & & - & \frac{1}{3}\cdot 0 & + & y_2 & & & = & 5 \\
\frac{1}{3}\cdot 0 & & & - & \frac{2}{3}\cdot 0 & & & - & z & = & -10
\end{array}
$$

よって、この時、$(x_1, y_1) = (0, 0)$ は基底可能解であり、$z = 10$ よって、係数を比較して、正の勾配となるのは $x_1$ の時のみであるため、$x_1$ を選び、$y_1 = 0$ として $x_1$ がどこまで動かせるか調べる。

$$
\begin{array}{lllllllllll}
\frac{1}{3}x_1 & + & x_2 & + & \frac{1}{3}\cdot 0 & & & & & = & 5 \\
\frac{5}{3}x_1 & & & - & \frac{1}{3}\cdot 0 & + & y_2 & & & = & 5 \\
\frac{1}{3}x_1 & & & - & \frac{2}{3}\cdot 0 & & & - & z & = & -10
\end{array}
$$

$$
\Longrightarrow
\begin{array}{lllllllllll}
\frac{1}{3}x_1 & + & 0 & + & \frac{1}{3}\cdot 0 & & & & & \leq & 5 \\
\frac{5}{3}x_1 & & & - & \frac{1}{3}\cdot 0 & + & 0 & & & \leq & 5 \\
\frac{1}{3}x_1 & & & - & \frac{2}{3}\cdot 0 & & & - & z & = & -10
\end{array}
$$

$$
\Longrightarrow
\begin{array}{lllllllllll}
x_1 & + & 0 & + & 3\cdot 0 & & & & & \leq & 15 \\
x_1 & & & - & \frac{1}{5}\cdot 0 & + & \frac{3}{5}\cdot 0 & & & \leq & 3 \\
\frac{1}{3}x_1 & & & - & \frac{2}{3}\cdot 0 & & & - & z & = & -10
\end{array}
$$

この場合 $x_1 = 3$ まで動かせることを意味する。すなわち $x_1 - \frac{1}{5}\cdot 0 + \frac{3}{5}\cdot 0 \leq 3$ が制約として聞くことが分かるので、以下のような式になる。

$$
\begin{array}{lllllllllll}
x_1 & + & 3\cdot x_2 & + & y_1 & & & & & = & 15 \\
x_1 & & & - & \frac{1}{5}\cdot y_1 & + & \frac{3}{5}\cdot y_2 & & & = & 3 \\
\frac{1}{3}x_1 & & & - & \frac{2}{3}\cdot y_1 & & & - & z & = & -10
\end{array}
$$

この時、方程式 $x_1 - \frac{1}{5}\cdot y_1 + \frac{3}{5}\cdot y_2 = 3$ を使って、$x_1$ を基底変数から外す。

$$
\begin{array}{lllllllllll}
 & + & x_2 & + & \frac{2}{5}y_1 & - & \frac{1}{5}y_2 & & & = & 4 \\
x_1 & & & - & \frac{1}{5}\cdot y_1 & + & \frac{3}{5}\cdot y_2 & & & = & 3 \\
 & & & - & \frac{3}{5}\cdot y_1 & - & \frac{1}{5}y_2 & - & z & = & -11
\end{array}
$$

この時、基底変数は $y_1, y_2$ であり、$(y_1, y_2) = (0, 0)$ は基底可能解であり、$z = 11$。また、この時 $y_1$ 方向も $y_2$ 方向の勾配もどちらも負であるため、これ以上値が大きくなることはないので、これが最適値

となる。

通常は以下のように係数比較だけで行う。

$$
\begin{array}{ccccccccccc}
x_1 & + & 3x_2 & + & y_1 & & & & & = & 15 \\
2x_1 & + & x_2 & & & + & y_2 & & & = & 10 \\
-x_1 & - & 2x_2 & & & & & + & z & = & 0
\end{array}
$$

$$
\Longrightarrow \begin{array}{ccccc}
1 & 3 & 1 & 0 & 15 \\
2 & 1 & 0 & 1 & 10 \\
-1 & -2 & 0 & 0 & 0
\end{array}
$$

$$
\Longrightarrow \begin{array}{ccccc}
1/3 & 1 & 1/3 & 0 & 5 \\
5/3 & 0 & -1/3 & 1 & 5 \\
-1/3 & 0 & 2/3 & 0 & 10
\end{array}
$$

$$
\Longrightarrow \begin{array}{ccccc}
0 & 3 & 6/5 & -3/5 & 12 \\
1 & 0 & -1/5 & 3/5 & 3 \\
0 & 0 & 9/5 & 3/5 & 33
\end{array}
$$

$$
\Longrightarrow \begin{array}{ccccc}
0 & 1 & 2/5 & -1/5 & 4 \\
1 & 0 & -1/5 & 3/5 & 3 \\
0 & 0 & 3/5 & 1/5 & 11
\end{array}
$$

## 1.4　双対問題

正準形に直した際にシンプレックス法で解けない場合に双対問題に変形して解くことがある。双対問題とは以下のように変形することをいう。

$$
\begin{array}{ccccc}
x_1 & + & 2x_2 & \leq & 16 \\
2x_1 & + & x_2 & \leq & 14 \\
20x_1 & + & 30x_2 & = & z
\end{array}
$$
$$
x_1 \geq 0, x_2 \geq 0
$$

の条件を満たす $z$ の最大化する問題は以下の双対問題として考えられる。

$$
\begin{array}{ccccc}
\xi_1 x_1 & + & 2\xi_1 x_2 & \leq & 16\xi_1 \\
2\xi_2 x_1 & + & \xi_2 x_2 & \leq & 14\xi_2 \\
20x_1 & + & 30x_2 & = & z
\end{array}
$$
$$
x_1 \geq 0, x_2 \geq 0, \xi_1 \geq 0, \xi_2 \geq 0
$$

$$
\Longrightarrow \begin{array}{ccccccc}
\xi_1 x_1 & + & 2\xi_1 x_2 & \leq & 16\xi_1 & & \\
+ & & + & & + & = & \zeta \\
2\xi_2 x_1 & + & \xi_2 x_2 & \leq & 14\xi_2 & & \\
\geq & & \geq & & \geq & & \\
20x_1 & + & 30x_2 & = & z & &
\end{array}
$$
$$
x_1 \geq 0, x_2 \geq 0, \xi_1 \geq 0, \xi_2 \geq 0
$$

$$
\Longrightarrow \begin{array}{ccccc}
\xi_1 & 2\xi_1 & 16\xi_1 & & \\
+ & + & + & = & \zeta \\
2\xi_2 & \xi_2 & 14\xi_2 & & \\
\geq & \geq & & & \\
20 & 30 & & &
\end{array}
$$
$$
\xi_1 \geq 0, \xi_2 \geq 0
$$

の時に $\zeta$ を最小化する問題となる。

よってこれを解くと以下のようになる。

$$
\begin{array}{ccccccccccc}
-\xi_1 & - & 2\xi_2 & + & y_1 & & & & & = & -20 \\
-2\xi_1 & - & \xi_2 & & & + & y_2 & & & = & -30 \\
-16\xi_1 & - & 14\xi_2 & & & & & + & \zeta & = & 0
\end{array}
$$

ここでこの時、初期解 $\xi_1 = \xi_2 = 0$ は制約条件を満たさないため、拡大された標準問題への変換を行い、それを解く。

最小化 $w = u_1 + u_2$

本来の目的関数$\zeta = 16\xi_1 + 14\xi_2$

$$
\text{制約条件}\left\{\begin{array}{rcrcrcrcrcrcr}
-\xi_1 & - & 2\xi_2 & + & y_1 & & & - & u_1 & & & = & -20 \\
-2\xi_1 & - & \xi_2 & & & + & y_2 & & & - & u_2 & = & -30
\end{array}\right.
$$

よって、この問題でシンプレックス法で解くことを考える。

$$
\begin{array}{rcrcrcrcrcrcrcr}
\xi_1 & + & 2\xi_2 & - & y_1 & & & & & + & u_1 & & & = & 20 \\
2\xi_1 & + & \xi_2 & & & - & y_2 & & & & & + & u_2 & = & 30 \\
-16\xi_1 & - & 14\xi_2 & & & & & + & \zeta & & & & & = & 0 \\
 & & & & & & & & w & - & u_1 & - & u_2 & = & ?
\end{array}
$$

ここで初期解は $\xi_1 = \xi_2 = y_1 = y_2 = 0, u_1 = 20, u_2 = 30$ であるが、これは目的関数 $w$ が非基底変数 (0 となっている変数) で表現されてないので掃き出し法により式変形する。

$$
\begin{array}{rcrcrcrcrcrcrcr}
\xi_1 & + & 2\xi_2 & - & y_1 & & & & & + & u_1 & & & = & 20 \\
2\xi_1 & + & \xi_2 & & & - & y_2 & & & & & + & u_2 & = & 30 \\
-16\xi_1 & - & 14\xi_2 & & & & & + & \zeta & & & & & = & 0 \\
3\xi_1 & + & 3\xi_2 & - & y_1 & - & y_2 & + & w & & & & & = & 50
\end{array}
$$

よって、ここから普通のシンプレックス法を用いる。$w$ を小さくするような寄与をする基底変数を選択していく。

$$
\Longrightarrow \begin{array}{cccccccc}
1 & 2 & -1 & 0 & 1 & 0 & 0 & 20 \\
2 & 1 & 0 & -1 & 0 & 1 & 0 & 30 \\
-16 & -14 & 0 & 0 & 0 & 0 & 1 & 0 \\
3 & 3 & -1 & -1 & 0 & 0 & 0 & 50
\end{array}
$$

$$
\Longrightarrow \begin{array}{cccccccc}
1 & 2 & -1 & 0 & 1 & 0 & 0 & 20 \\
1 & 1/2 & 0 & -1/2 & 0 & 1/2 & 0 & 15 \\
-16 & -14 & 0 & 0 & 0 & 0 & 1 & 0 \\
3 & 3 & -1 & -1 & 0 & 0 & 0 & 50
\end{array}
$$

$$
\Longrightarrow \begin{array}{cccccccc}
0 & 3/2 & -1 & 1/2 & 1 & -1/2 & 0 & 5 \\
1 & 1/2 & 0 & -1/2 & 0 & 1/2 & 0 & 15 \\
0 & -6 & 0 & -8 & 0 & 8 & 1 & 240 \\
0 & 3/2 & -1 & 1/2 & 0 & -3/2 & 0 & 5
\end{array}
$$

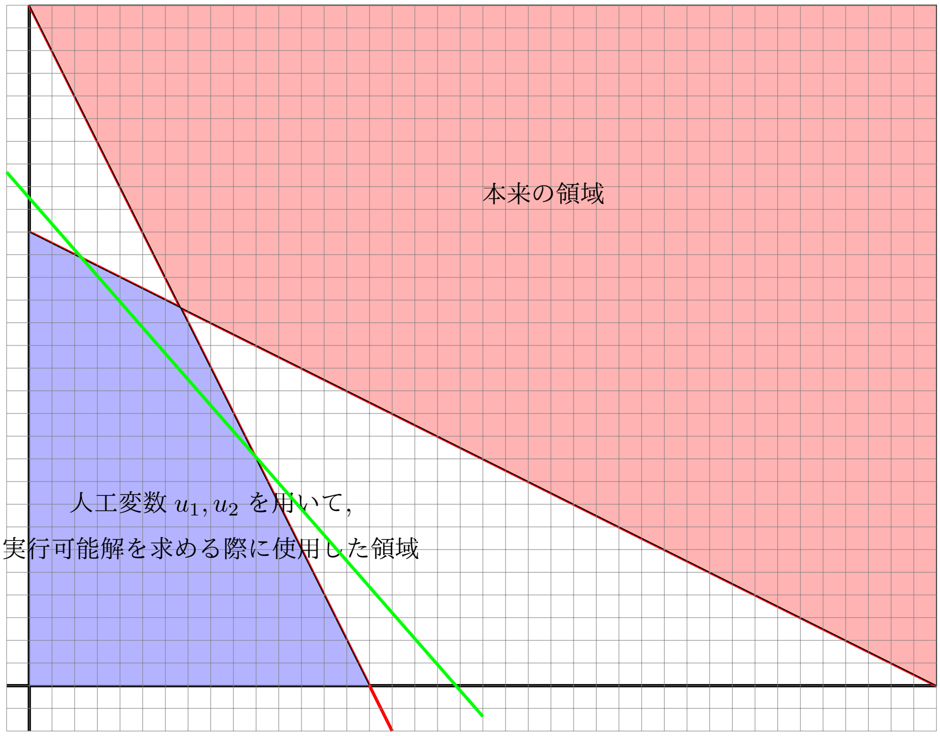
$$
\Longrightarrow \begin{array}{cccccccc}
0 & 1 & -2/3 & 1/3 & 2/3 & -1/3 & 0 & 10/3 \\
2 & 1 & 0 & -1 & 0 & 1 & 0 & 30 \\
0 & -6 & 0 & -8 & 0 & 8 & 1 & 240 \\
0 & 3/2 & -1 & 1/2 & 0 & -3/2 & 0 & 5
\end{array}
$$

$$
\Longrightarrow \begin{array}{cccccccc}
0 & 1 & -2/3 & 1/3 & 2/3 & -1/3 & 0 & 10/3 \\
2 & 0 & 2/3 & -4/3 & -2/3 & 4/3 & 0 & 80/3 \\
0 & 0 & -4 & -6 & 4 & 6 & 1 & 260 \\
0 & 0 & 0 & 0 & -1 & -1 & 0 & 0
\end{array}
$$

よって、非基底変数 $y_1 = y_2 = u_1 = u_2 = 0$ であるので、$\xi_1 = 40/3, \xi_2 = 10/3, \zeta = 260$ となり、ようやく実行可能解が求まったので、$u_1, u_2$ を排除して計算する。

$$
\begin{array}{rrcrcrcrcr}
 & \xi_2 & - & \frac{2}{3}y_1 & + & \frac{1}{3}y_2 & & & = & \frac{10}{3} \\
2\xi_1 & & + & \frac{2}{3}y_1 & - & \frac{4}{3}y_2 & & & = & \frac{80}{3} \\
 & & - & 4y_1 & - & 6y_2 & + & \zeta & = & 260
\end{array}
$$

よって、ここから $\zeta$ を最小化させるようにシンプレックス法を用いるのだが、この時すでに最適解となっているため、これで終了となる。よって、最適解は $\zeta = 260$ となる。




2　生成モデル

パターン認識には以下の 2 種類が存在する。

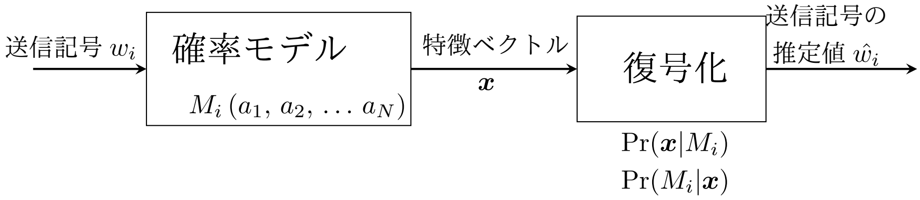
1. 識別モデルに基づくパターン認識
   特徴空間に対し、各クラスの境界を求めておき、クラス未知のデータの特徴ベクトルが特徴空間上のどこに位置するかを調べることでデータクラスを定める方法。
   $\Longrightarrow$ クラス未知のデータの特徴ベクトルに対してダイレクトにクラスを判定する関数が存在する。
2. 生成モデルに基づくパターン認識
   クラスごとに特徴ベクトルの出力確率を与える確率モデルを用意し、クラス未知のデータの特徴ベクトルがそれぞれの確率モデルから出力する確率を調べることでクラスを定める方法。
   $\Longrightarrow$ 各クラスの確率モデルを与える出力確率を比較することによりクラス境界が定まるため、識別モデルよりも 1 ステップ多いが、個別の確率のモデルをモデル化しやすく、ほかの特徴空間の識別に対しても知見を与えることができる利点も存在する。

2.1　識別モデル

識別モデルには SVM などがあるが、今回は割愛する。

2.2　生成モデル

生成モデルによるパターン認識は以下のような図で表されるようなモデルである。




送信記号が $w_i$ から、特徴ベクトルである $\boldsymbol{x}$ への写像を確率モデル $M$ で表し、それにより、抽出された特徴ベクトルから、送信記号 $\hat{w}_i$ が送信された確率である事後確率 $\Pr(\hat{w}_i|\boldsymbol{x})$ を求めることにより、復号化し、その記号から境界線を定めることにより、クラスを推定し、送信記号が、何であったか推定す

るというものである。

2.2.1　ベイズ決定則

この部分は授業では飛ばされたが、この部分が理解できないと次に進めないと考えられるため、記述する。

$\{w_i\}$ をクラス集合、$\{\alpha_i\}$ を取りうる行動の集合、$\lambda(\alpha_i|w_j)$ を $x$ のクラスが実際は $w_j$ であるときに行動 $\alpha_i$ を取ることの損失とする。

ここで、$\alpha_i$ はデータ $x$ のクラスを $w_i$ と判断することを意味し、$\alpha_0$ はデータ $x$ のクラスは判定不能とすることを意味する。

この時 $R(\alpha_i|x) = \sum_j \lambda(\alpha_i|w_j)P(w_j|x)$ を条件付きリスクと呼び、この値が最小となる行動 $\alpha$ を選択する決定則をベイズ決定則と呼び、以下の式で表される。

$$\alpha = \operatorname*{argmin}_i R(\alpha_i|x) = \operatorname*{argmin}_{\alpha_i} \sum_j \lambda(\alpha_i|w_j)\mathrm{Pr}(w_j|x) \tag{2.1}$$

この時、損失 $\lambda(\alpha_i|w_j)$ を具体的に以下のように定義する。

$$\lambda(\alpha_i|w_j) = \begin{cases} 0 & if\ i = j \\ 1 & if\ i \neq j \end{cases}$$

正解を与える行動には 0 を不正解には 1 を与えることを意味する

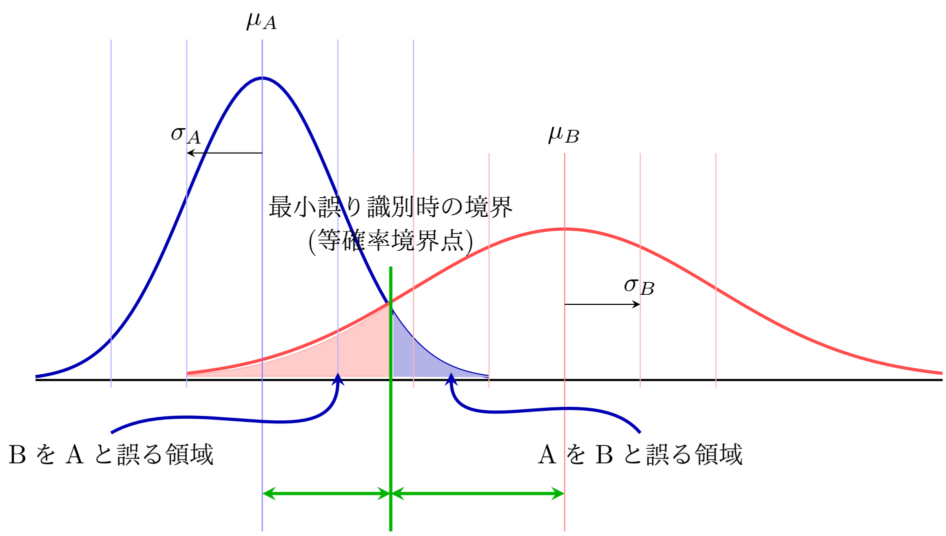
この時、条件付きリスクは以下のように表される。

$$\begin{aligned} R(\alpha_i|x) &= \sum_j \lambda(\alpha_i|w_j)P(w_j|x) \\ &= \sum_{j\neq i} \mathrm{Pr}(w_j|x) \\ &= 1 - \mathrm{Pr}(wj|x) \end{aligned}$$

となるため、ベイズ決定則は式 (2.1) より、以下のように最小誤り率を与える決定則となる。

$$\begin{aligned} \alpha &= \operatorname*{argmin}_i R(\alpha_i|x) \\ &= \operatorname*{argmin}_i 1 - \mathrm{P}(w_i|x) \\ &= \operatorname*{argmax}_i \mathrm{P}(w_i|x) \end{aligned} \tag{2.2}$$

従って、誤り率最小化は事後確率を最大化するクラスを選ぶことで実現できる。尚、誤り率は以下のようなもののこと表す。



この図の様に最小誤り率は 2 クラスの場合は二つの分布の交点を通る直線を境界線としたときに誤り率が最小になる。

式 (2.2) の時、$w_\alpha$ が送信されたと判断する復号法は最大事後確率復号と呼ばれ、これによりクラス識別が可能になるが、一般にすべての $\Pr(w_i|x)$ をすべての $x$ に対して用意することは不可能であるため、最尤基準による復号を考える。

2.3　最尤基準による復号

Bayes の定理により以下が成り立つ。

$$\Pr(w_i|x) = \frac{\Pr(x|w_i)\Pr(w_i)}{\Pr(x)}$$

よって、この時、$\Pr(x)$ はクラスに無関係。よって、以下のように変形できる

$$\operatorname{argmax}\Pr(w_i|x) = \operatorname{argmax}\Pr(x|w_i)\Pr(w_i)$$

これはクラスごとにデータを集め、分布を調べれば $\Pr(x|w_i)$ は求まる。特に $\Pr(w_i)$ はクラスによって変わらないとすれば以下のようになる。

$$\operatorname{argmax}\Pr(w_i|x) = \operatorname{argmax}\Pr(x|w_i)$$

この求め方を最尤復号という。

3　最尤推定・EM アルゴリズム

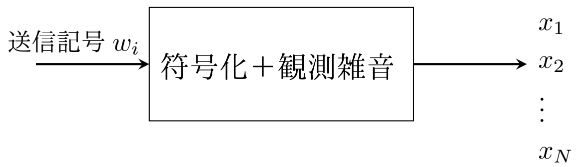
2.3 より、$\Pr(x|w_i)$ が求まればパターン認識ができるがこの確率をどのように求めるのかが問題となる。

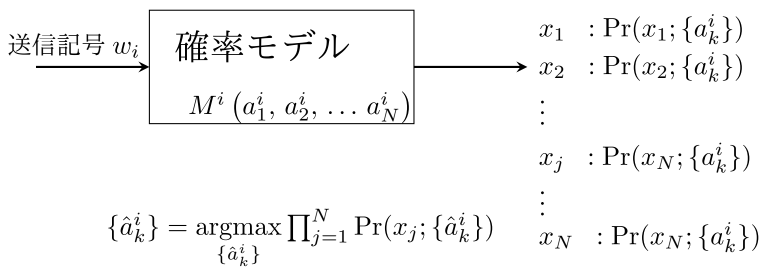
この時、常套手段として用いられるのは $w_i$ から $x$ への写像を確率モデル $M_i(a_1, a_2, \ldots, a_N)$ で表し、この確率モデルのパラメータを大量の学習データを用いて推定する。

3.1　最尤推定

カテゴリ $w_i$ のデータを生成して、そのパターン $x_1, x_2, \ldots, x_N$ を収集する。

その後、各々のデータ $x_j$ がカテゴリ $w_i$ の確率モデルから生起する確率が高くなるようにモデルのパラメタ $\{\hat{a}_k^i\}$ を調整する。



### 3.2　最尤推定の例

確率分布が正規分布であり、以下のようなモデルの時、尤度は式 (3.1) のようになる。

$$
\begin{aligned}
M &= \{\mu, \sigma^2\} \\
\Pr(x;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \\
\Longrightarrow P &= \prod_{j=1}^{N}\Pr(x_j;\mu,\sigma^2) \\
\log P &= \sum_{j=1}^{N}\log\Pr(x_j;\mu,\sigma^2) \\
&= -\frac{N}{2}\log 2\pi - \frac{N}{2}\log\sigma^2 - \sum_{j=1}^{N}\frac{1}{2}\frac{(x_j-\mu)^2}{\sigma^2} \qquad (3.1)
\end{aligned}
$$

ここで、$\log P$ を各パラメタで偏微分して 0 とおく

$$
\begin{aligned}
&\frac{\partial \log P}{\partial \mu} = -\sum_{j=1}^{N}\frac{x_j-\mu}{\sigma^2} = 0 \\
&\mu = \frac{1}{N}\sum_{j=1}^{N}x_j \\
&\frac{\partial \log P}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \sum_{j=1}^{N}\frac{(x_j-\mu)^2}{2\sigma^4} = 0 \\
&\sigma^2 = \frac{1}{N}\sum_{j=1}^{N}(x_j-\mu)^2
\end{aligned}
$$

確率分布が混合正規分布の時、

$$
\begin{aligned}
M &= \{\mu_1,\sigma_1^2,m_1,\mu_2,\sigma_2^2,m_2,\mu_3,\sigma_3^2,m_3\} \\
\mathrm{P}\left(x;\mu_k,\sigma_k^2\right) &= \frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\left(-\frac{1}{2}\frac{(x-\mu_k)^2}{\sigma_k^2}\right) \\
\Pr(x;M) &= \sum_{k=1}^{3}m_k\mathrm{P}\left(x;\mu_k,\sigma_k^2\right), \sum_{k=1}^{3}m_k = 1
\end{aligned}
$$

この時はパラメタ毎に偏微分して 0 とおいても解けないので、EM アルゴリズムを用いて推定する。

### 3.3　EM アルゴリズム

モデル $M$ の出力 $x$ の振る舞いを決めるのに観測できないデータ（潜在変数）$y$ が関与する枠組みにおいて観測できるデータ $x$ のみを用いて尤度 $\Pr(x|M)$ を最大化するパラメタ $M$ を求める問題を解く。

$$
\underset{M}{max}\prod\Pr(x|M) \qquad (3.2)
$$

観測不能な $y$ が存在するなら、式 (3.2) は簡単には実行できないが、$y$ の出力を仮定すると $(x,y)$ の同時確率は計算できるため、式 (3.3) は容易に実行できる。

$$
\underset{M}{max}\prod\Pr((x,y)|M) \qquad (3.3)
$$

よって、式 (3.2) の問題を (3.3) に置き換えて解く。このことを実現するには $x$ とモデルの初期推定値 $M$ が与えられた条件における完全データ $(x,y)$ に対する $\log\Pr(x|M')$ の期待値を最大化する問題を考

える。

$$\begin{aligned}
&\underset{M'}{max}\, E\left[\log \Pr(x|M')|x,M\right] \\
\text{ここで } E\left[f(x,y)|x,M\right] &= \sum f(x,y_i)\Pr(y_i|x,M) \\
\text{この時、}E\left[\log \Pr(x|M')|x,M\right] &= \sum \Pr(y_i|x,M)\log \Pr(x|M') \\
&= \log \Pr(x|M') \\
&= L(x,M')
\end{aligned}$$

よってこの時、以下のことが成り立つ。

$$\begin{aligned}
L(x,M') &= E\left[\log \Pr(y_i|x,M')|x,M\right] \\
&= \sum \Pr(y_i|x,M)\log \Pr(x|M') \\
&= \sum \Pr(y_i|x,M)\log \frac{\Pr(y_j|x,M')\Pr(x|M')}{\Pr(y_j|x,M')} \\
&= \sum \Pr(y_i|x,M)\log \frac{\Pr(x,y_j|M')}{\Pr(y_j|x,M')} \\
&= \sum \Pr(y_i|x,M)\log \Pr(x,y_j|M') - \sum \Pr(y_i|x,M)\log \Pr(y_j|x,M') \\
\text{ここで}& \\
Q(M,M') &= E\left[\log \Pr(x,y|M')|x,M\right] \\
&= \sum \Pr(y_i|x,M)\log \Pr(x,y_j|M') \\
H(M,M') &= E\left[\log \Pr(y|x,M')|x,M\right] \\
&= \sum \Pr(y_i|x,M)\log \Pr(y_j|x,M')
\end{aligned}$$

とおくと、次式が成り立つ

$$L(x,M') = Q(M,M') - H(M,M')$$

$Q(M,M')$ はモデルの初期値と学習データで条件づけられた完全データの対数尤度の期待値である。$H(M,M')-H(M,M)$ はダイバージェンスであるため、

$$\begin{aligned}
H(M,M') &\leq H(M,M) \\
\because H(M,M')-H(M,M) &= \sum \Pr(y_i|x,M)\left\{\log \Pr(y_j|x,M') - \log \Pr(y_j|x,M)\right\} \\
&= \sum \Pr(y_i|x,M)\log \frac{\Pr(y_j|x,M')}{\Pr(y_j|x,M)} \\
&\leq \sum \Pr(y_i|x,M)\left(\frac{\Pr(y_j|x,M')}{\Pr(y_j|x,M)} - 1\right) \\
&\leq \sum \left(\Pr(y_j|x,M) - \Pr(y_j|x,M')\right)
\end{aligned}$$

よって、$H(M,M')$ はパラメタ変化に対して単調減少関数である。よって、$Q(M,M')$ を最大化するように $M'$ を変化させれば $L(x,M')$ はは必ず増大する。従って、$Q(M,M')$ を最大化する $M'$ を求めては $M=M'$ とし、また、$Q(M,M')$ を最大化する $M'$ を求めるという処理を繰り返せば $M'$ は $L(x,M')$ の極大値を与える $M'$ に収束する。$\rightarrow$ 不完全データの対数尤度 $\log \Pr(x|M)$ の最大化問題が完全データ $\log \Pr(x,y|M)$ の最大化する問題に置換された。よって、具体的なアルゴリズムとしては以下のようになる。

1. 適当な初期モデル $M$ を選ぶ
2. 与えられた学習データ x を用いて対数尤度 $\log \mathrm{P}$ の期待値 $Q(M,M')$ を最大化する $M'$ を求める。

   2-1 $\Pr(y_j|x_j,M)$ を求める。

   2-2 $Q(M,M') = \sum_{j=1}^{12} \Pr(y_j|x_j,M)\log \Pr(x_j,y_j|M')$ を最大化する $M'$ を求める

3. 収束していれば $M'$ を答えとして終了する
4. $M = M'$ とおいて 2 に戻る。

### 3.4　例題

３つの壺があって、それぞれの壺には正規分布に従ってボールが入っているものとする。適当な確率で壺を選んでその壺からボールを取り出し、そのボールの重さを報告してまた壺に戻すという操作を繰り返したところ、図 3.1 のような報告になった。各壺にあるボールの重さの分布を EM アルゴリズムで求めよ。

| 重さ: | 8 | 3 | 6 | 9 | 7 | 6 | 2 | 6 | 4 | 4 | 5 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

図 3.1　報告結果

ただし、パラメタの初期値は式 (3.4), (3.5), (3.6), のとおりとする。

$$\mathrm{P(A)} = \frac{1}{3}, \mathrm{M_A} = \left\{\mu_A,\ {\sigma_A}^2\right\} = \{8, 1.0\} \tag{3.4}$$

$$\mathrm{P(B)} = \frac{1}{3}, \mathrm{M_B} = \left\{\mu_B,\ {\sigma_B}^2\right\} = \{3, 1.0\} \tag{3.5}$$

$$\mathrm{P(C)} = \frac{1}{3}, \mathrm{M_C} = \left\{\mu_C,\ {\sigma_C}^2\right\} = \{6, 1.0\} \tag{3.6}$$

### 3.5　解答

この問題は正規分布が 3 つ重なった確率分布 $M = \{\mu_1, \sigma_1^2, m_1, \mu_2, \sigma_2^2, m_2, \mu_3, \sigma_3^2, m_3\}$ の混合正規分布であるため、EM アルゴリズムを使用できる。よって、EM アルゴリズムに基づき、書いたコードの実行結果は図 3.1 のようになる。

```
          P()                        mu                   sigma
A : 0.16286318721123902 | 8.496670931853629 | 0.2798250295558986
B : 0.41417380589888836 | 3.6340475605911426 | 1.2014584126825552
C : 0.42296300688987265 | 6.3703233422503684 | 0.35114964150890327
尤度 : 3.234005932269144E-11
```

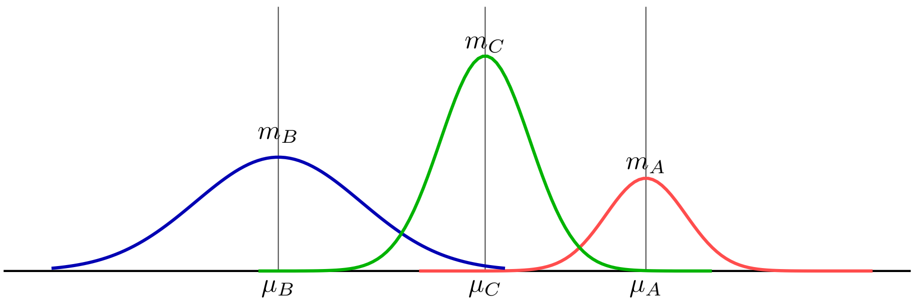
図 3.1　実行結果

従って、求める分布は以下のようになる。

$$\mathrm{P(A)} = 0.1629,\ \mathrm{M_A} = \left\{\mu_A,\ \sigma_A^2\right\} = \{8.497,\ 0.2798\} \tag{3.7}$$

$$\mathrm{P(B)} = 0.4142,\ \mathrm{M_B} = \left\{\mu_B,\ \sigma_B^2\right\} = \{3.634,\ 1.201\} \tag{3.8}$$

$$\mathrm{P(C)} = 0.4230,\ \mathrm{M_C} = \left\{\mu_C,\ \sigma_C^2\right\} = \{6.370,\ 0.3511\} \tag{3.9}$$



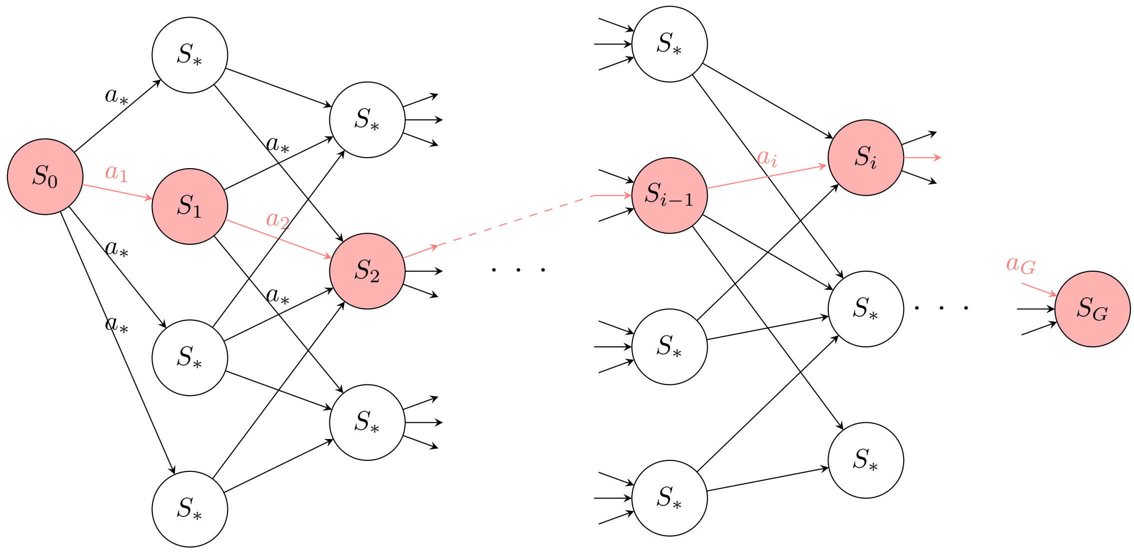

## 4　動的計画法

動的計画法とは逐次決定過程の最適化問題を解く方法のことをいう。

### 4.1　逐次決定過程

逐次決定過程とは過程の各状態において意思決定者が何らかの決定を選択すると状態遷移が起こり、対応する利得（あるいは費用）が発生すると考える数学モデル。

過程から発生する総利得（費用）を最適化する政策（すなわち決定の系列）を求めようとすることが、逐次決定過程の最適化問題。



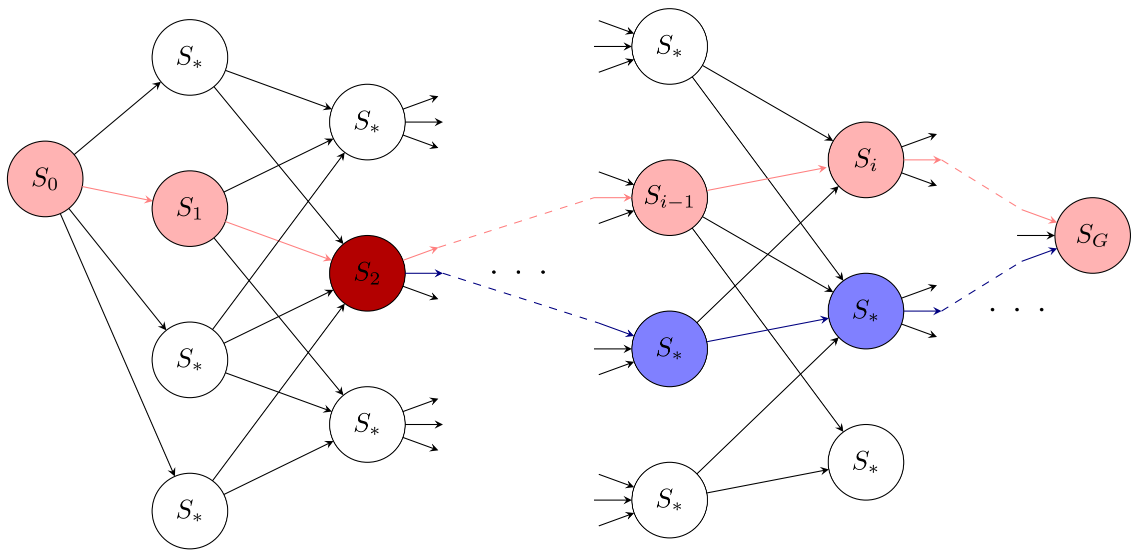

### 4.2　ベルマンの最適性の原理

◎ 最適政策では初期状態、初期の決定が何であろうと以後の政策は最初の遷移から生じた状態に関して適切でなければならない。

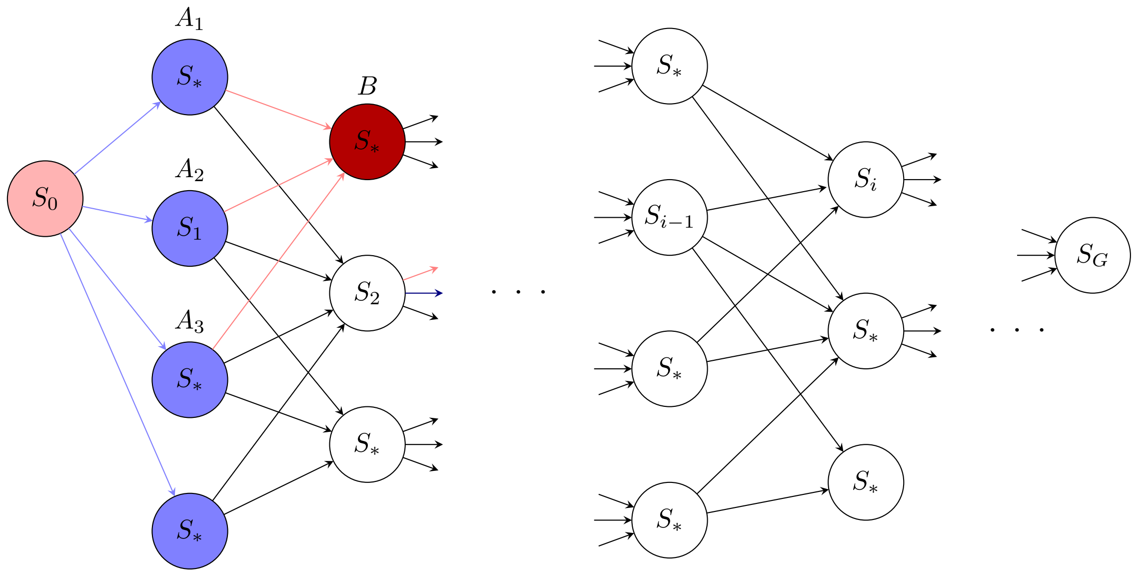
◎ ある期間を通じての最適問題の解として最適政策は元々の問題を部分区間に区切った部分最適問題の解を一部として持つ。

最適政策上にある状態を考えると、$S_2$ から $SG$ への最適政策も赤のパスに一致する。

この時、下のように青のパスが最適政策とする場合を考える。




この時、仮に青のパスが $S_2$ から $S_G$ までの最適政策なら全体の最適政策も $S_0 \rightarrow S_2 \rightarrow S_G$ となって仮定に反する。




この時、$A_i$ に続く状態 B への $A_i$ 経由の最適政策のコストは $A_i$ への最適コスト（求まっているとする。）と $A_i$ から B へのコストの和となる。

$$\mathrm{Optimalcost(S_0 \to A_i \to B) = Optimalcost(S_0 \to A_i) + Cost(A_i \to B)}$$

よって、$S_0$ から $B$ への最適コストは以下のようになる。

$$\mathrm{Optimalcost(S_0 \to B) = max[Optimalcost(S_0 \to A_i \to B)]}$$

## 5　HMM

HMM には以下の 2 種類が存在する。

1. エルゴディック HMM：既約で（どこへでも状態遷移できる）で周期的でない HMM
2. left-to-right HMM:帰還ループを持たない（一度ある状態を抜けたなら、二度とその状態には戻ってこない）HMM

音響モデルとしては確率分布を切り替えながら時系列を生成するものとモデル化できるため、left-to-right HMM が用いられる。
また、音声の場合は必ず一定のものではなく、動的で連続的な特徴ベクトルの集合であるため、確率分布から確率分布への遷移の軌跡を表すことができるという点で left-to-right HMM は用いられている。
基本的にデータが与えられて、このモデルに沿った HMM が与えられている際にそれぞれのデータがどの確率分布から出てきたかの割り当ては対応付けの数だけありうるため、組み合わせ爆発を起こしてしまう。
そのため、経路問題のような感じで効率的に確率計算を行うことを考える。その時に用いられるアルゴリズムを 2 つ紹介する。

### 5.1　フォワードアルゴリズム

HMM モデルにおけるデータ $x_1$ $x_T$ までが出力される確率を表す。そのため、以下のような漸化式を用いて求められる。

$$\alpha(t, j) = \sum_k \alpha(t-1, k) a_{kj} b_j(x_t)$$

$a_{kj}$ は状態 $k$ から状態 $j$ への遷移確率であり、$b_j(x_t)$ は状態 $j$ において $x_t$ が出現する確率であり、$\alpha(t, j)$ は時刻 $t$ までに $x_1$ $x_t$ を出力して状態 $j$ に至る確率である。よって、最終的な確率は以下のようになる。

$$\Pr(x_1, ...x_T) = \sum_k \alpha(T, k)$$

### 5.2 ビタビアルゴリズム

このアルゴリズムは最適な状態遷移が生じた場合の出力確率を求めるアルゴリズムである。そのため、総和を求めるのではなく、最適な一つのパスのみを求めるものである。以下の式で表される。

$$\alpha(t,j) = \max_k \alpha(t-1,k) a_{kj} b_j(x_t)$$

よって、最終的な確率は以下のように表される。

$$\Pr(x_1,...,x_T) = \max_k \alpha(T,k)$$

## 6 参考文献


[1] "F2-viterbialgorithm.pdf." https://ocw.u-tokyo.ac.jp/lecture_files/engin_01/6/notes/ja/F2-ViterbiAlgorithm.pdf. (Accessed on 01/21/2020).

[2] "行列式の基本的な性質と公式 - 理数アラカルト -." https://risalc.info/src/determinant-formulas.html. (Accessed on 01/21/2020).

[3] "「ベクトルで微分・行列で微分」公式まとめ - qiita." https://qiita.com/AnchorBlues/items/8fe2483a3a72676eb96d. (Accessed on 01/21/2020).

[4] "線形計画法の双対定理の意味と嬉しさ | 高校数学の美しい物語." https://mathtrain.jp/primaldual. (Accessed on 01/21/2020).

[5] "twophase2.dvi." http://www.fujilab.dnj.ynu.ac.jp/lecture/system4.pdf. (Accessed on 01/21/2020).

[6] "dual2002.pdf." http://www.bunkyo.ac.jp/~nemoto/lecture/mathpro/2002/dual2002.pdf. (Accessed on 01/21/2020).

[7] "exercise1.dvi." http://www.fujilab.dnj.ynu.ac.jp/lecture/system2.pdf. (Accessed on 01/21/2020).