Pythonで動かして学ぶ！あたらしい機械学習の教科書 第2版, 伊藤真(翔泳社).

$$\left(\frac{g(x)}{f(x)}\right)' = \frac{g'(x)f(x) - g(x)f'(x)}{f(x)^2} \tag{4-123}$$

ここで、$f'(x) = \partial f/\partial x_1$、$g'(x) = \partial g/\partial x_1$と、$x_1$による偏微分を考えます。

$$f'(x) = \frac{\partial}{\partial x_1} f(x) = \exp(x_1)$$
$$g'(x) = \frac{\partial}{\partial x_1} \exp(x_0) = 0$$

よって式 4-128 のようになります。

$$\begin{aligned}\frac{\partial y_0}{\partial x_1} &= \frac{g'(x)f(x) - g(x)f'(x)}{f(x)^2} = \frac{-\exp(x_0)\exp(x_1)}{u^2} \\ &= -\frac{\exp(x_0)}{u} \cdot \frac{\exp(x_1)}{u}\end{aligned} \tag{4-128}$$

ここで、$y_0 = \exp(x_0)/u$、$y_1 = \exp(x_1)/u$であったことを使うと、式 4-129 が得られます。

$$\frac{\partial y_0}{\partial x_1} = -y_0 y_1 \tag{4-129}$$

式 4-126 と式 4-129 をまとめて、式 4-130 のように表すこともできます。

$$\frac{\partial y_j}{\partial x_i} = y_j(I_{ij} - y_i) \tag{4-130}$$

ここで、I_{ij}は、$i = j$のときに 1、$i \neq j$のときに 0 となる関数です。I_{ij}は、δ_{ij}とも表され、クロネッカーのデルタと呼ばれています。

4.7.7 ソフトマックス関数とシグモイド関数

それにしても、ソフトマックス関数とシグモイド関数は似ています。これは、どういう関係なのでしょうか？　ここで、その関係を考えてみましょう。2 変数の場合のソフトマックス関数は、式 4-131 のようになります。

$$y = \frac{e^{x_0}}{e^{x_0} + e^{x_1}} \tag{4-131}$$

分母分子にe^{-x_0}を掛けて整理すると、$e^a e^{-b} = e^{a-b}$という公式を使って、式 4-132 を得ることができます。

$$y = \frac{e^{x_0}e^{-x_0}}{e^{x_0}e^{-x_0} + e^{x_1}e^{-x_0}} = \frac{e^{x_0-x_0}}{e^{x_0-x_0} + e^{x_1-x_0}} = \frac{1}{1 + e^{-(x_0-x_1)}} \tag{4-132}$$

ここで、$x = x_0 - x_1$とおけば、式 4-133 のようにシグモイド関数になりました。

$$y = \frac{1}{1 + e^{-x}} \tag{4-133}$$

つまり、2 変数のソフトマックス関数の入力x_0、x_1を、その差$x = x_0 - x_1$で表したものがシグモイド関数なのです。シグモイド関数を多変数に拡張したものがソフトマックス関数だとも言えます。

4.7.8 ガウス関数

ガウス関数は、式 4-134 のような関数です。

$$y = \exp(-x^2) \tag{4-134}$$

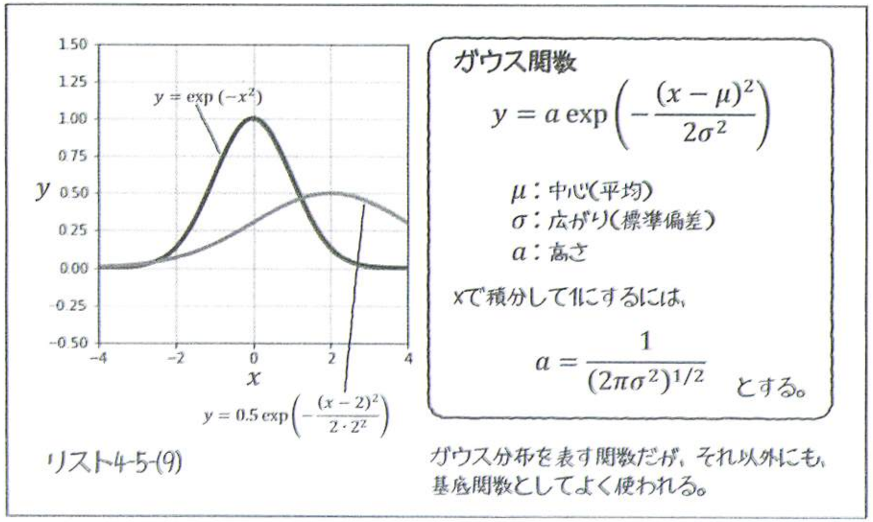


図 4.35：ガウス関数

図 4.35 左黒線のようにガウス関数は、$x = 0$を中心とした、釣鐘のような形をしています。ガウス関数は曲線を近似する基底関数として第 5 章で登場します。

この関数の中心（平均）をμで表し、広がりの大きさ（標準偏差）をσ、高さをaで調節できる形にすると、式 4-135 のようになります（図 4.35 左灰色線）。

$$y = a\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{4-135}$$

早速、グラフで描いてみましょう（リスト 4-5-(9)、図 4.35 左）。

In

```
# リスト 4-5-(9)
# ガウス関数 ----------
def gauss(mu, sigma, a):
    # 式 4-135
    y = a * np.exp(-((x - mu) ** 2) / (2 * sigma ** 2))
    return y


# グラフ描画 ----------
x = np.linspace(-4, 4, 100)
```



```
plt.figure(figsize=(4, 4))
plt.plot(x, gauss(0, 1, 1), "black", linewidth=3)
plt.plot(x, gauss(2, 2, 0.5), "gray", linewidth=3)
plt.xlim(-4, 4)
plt.ylim(-0.5, 1.5)
plt.grid()
plt.show()
```

Out

```
# 実行結果は図 4.35 を参照
```

ガウス関数で確率分布を表すことがありますが、その場合には、x に関した積分が 1 になるように、式 4-135 の a を、式 4-136 のようにします。

$$a = \frac{1}{(2\pi\sigma^2)^{1/2}} \tag{4-136}$$

4.7.9 2 次元のガウス関数

ガウス関数を 2 次元に拡張することができます。2 次元のガウス関数は、第 9 章の混合ガウスモデルで出てきます。

入力を 2 次元ベクトル $\mathbf{x} = [x_0, x_1]^T$としたとき、ガウス関数の基本形は、式 4-137 のようになります。

$$y = \exp\left\{-(x_0^2 + x_1^2)\right\} \tag{4-137}$$

グラフで表すと図 4.36 のように、原点を中心とした同心円状の釣鐘のような形になります。

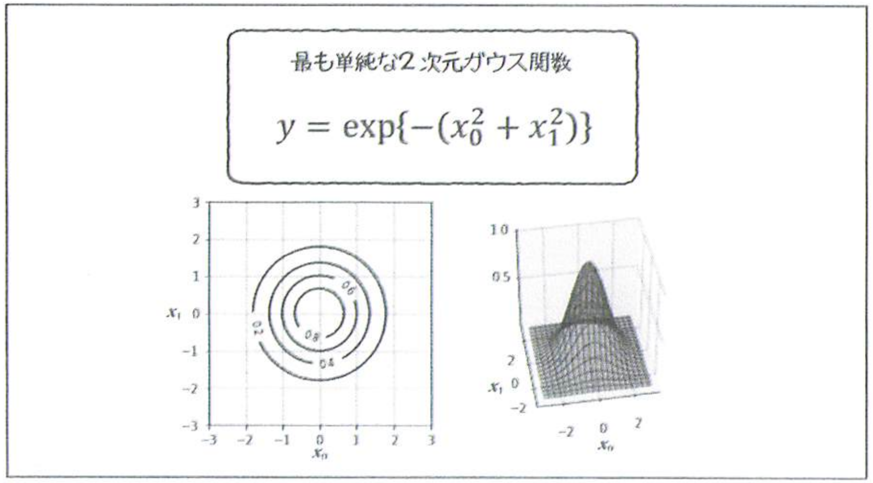


図4.36：単純な2次元ガウス関数

これを基本形として、中心を移動させたり、長細くしたりするために、いくつかのパラメータを加えた形が、式4-138です。

$$y = a \cdot \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{4-138}$$

こうなると、expの中にベクトルや行列が入っていたりするので、少し面食らうかもしれませんが、大丈夫です。1つ1つ解説していきます。

まず、関数の形を表すパラメータは$\boldsymbol{\mu}$と$\boldsymbol{\Sigma}$です。$\boldsymbol{\mu}$は、平均ベクトル（中心ベクトル）と呼ばれるパラメータで、関数の広がりの中心を表します（式4-139）。

$$\boldsymbol{\mu} = [\mu_0 \ \mu_1]^{\mathrm{T}} \tag{4-139}$$

$\boldsymbol{\Sigma}$は、共分散行列と呼ばれるもので、以下のように表される2×2の行列です。

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \tag{4-140}$$

その行列要素のσ_0^2とσ_1^2には、正の数値を割り当てることができ、それぞれx_0方向とx_1方向の関数の広がりの大きさを調節します。σ_{01}には正から負の実数を割り当て、関数の広がり方向の傾きを調節します。正の数だと右上がりに傾いた楕円状に、

負の数だと左上がりに傾いた楕円状に広がった形になります（x_0を横軸、x_1を縦軸とした場合）。

式4-138のexpの中身には、ベクトルや行列が入っていますが、整理するとスカラー量になります。例えば、単純になるように$\boldsymbol{\mu} = [\mu_0 \ \mu_1]^{\mathrm{T}} = [0 \ 0]^{\mathrm{T}}$だとして、$(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$を計算してみると、式4-141のようになり、expの中身の実体は、x_0とx_1でできた2次式（2次形式）であることがわかります。

$$\begin{aligned} (\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= \mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x} \\ &= [x_0 \quad x_1] \cdot \frac{1}{\sigma_0^2\sigma_1^2 - \sigma_{01}^2}\begin{bmatrix} \sigma_1^2 & -\sigma_{01} \\ -\sigma_{01} & \sigma_0^2 \end{bmatrix}\begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \\ &= \frac{1}{\sigma_0^2\sigma_1^2 - \sigma_{01}^2}(\sigma_1^2 x_0^2 - 2\sigma_{01}x_0x_1 + \sigma_0^2 x_1^2) \end{aligned} \tag{4-141}$$

aは、関数の大きさをコントロールするパラメータとも考えられますが、2次元のガウス関数で確率分布を表す場合には、式4-142のようにセットします。

$$a = \frac{1}{2\pi}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \tag{4-142}$$

上記のようにすることで、入力空間での積分値が1となり、関数が確率分布を表すことができます。

式4-142の$|\boldsymbol{\Sigma}|$は、$\boldsymbol{\Sigma}$の行列式と言われる量で、2×2の行列の場合には、式4-143の公式で計算される量です。

$$|\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{4-143}$$

よって$|\boldsymbol{\Sigma}|$は、式4-144のように表せます。

$$|\boldsymbol{\Sigma}| = \sigma_0^2\sigma_1^2 - \sigma_{01}^2 \tag{4-144}$$

それでは、Pythonのプログラムで描いてみましょう。まず、ガウス関数をリスト4-6-(1)で定義します。

In

```
# リスト 4-6-(1)
%matplotlib inline
import numpy as np
```

```
import matplotlib.pyplot as plt


# ガウス関数 ----------
def gauss(x0, x1, mu, sigma):
    x = np.array([x0, x1])
    # 式4-142
    a = 1 / (2 * np.pi) * 1 / (np.linalg.det(sigma) ** (1 / 2))
    # 式4-138
    inv_sigma = np.linalg.inv(sigma)
    y = a * np.exp(
        (-1 / 2) * (x - mu).T @ inv_sigma @ (x - mu))
    return y
```

引数の入力データ **x0** と **x1** は、単一の数値(ベクトルや行列には対応していません)、**mu** は大きさ２のベクトル、**sigma** は２×２の行列です。ここに、適当な数値を代入して **gauss(x0, x1, mu, sigma)** をテストします（リスト 4-6-(2)）。

In

```
# リスト 4-6-(2)
x0, x1 = 2, 1
mu = np.array([1, 2])                # 平均ベクトル
sigma = np.array([[1, 0], [0, 1]])   # 共分散行列
y = gauss(x0, x1, mu, sigma)
print("y =", np.round(y, 6))
```

Out

```
y = 0.05855
```

上記の結果から、入力した値に対する関数の値が返ってくることが確かめられました。この関数を、等高線表示と 3D 表示させたものがリスト 4-6-(3) です。

In

```
# リスト 4-6-(3)
# パラメータ ----------
mu = np.array([1, 0.5])              # (A) 平均ベクトル
sigma = np.array([[2, 1], [1, 1]])   # (B) 共分散行列
x0_min, x0_max = -3, 3               # x0の計算範囲
```



```
x1_min, x1_max = -3, 3               # x1の計算範囲

# データ生成 ----------
x0_n, x1_n = 40, 40  # グラフ表示の解像度
x0 = np.linspace(x0_min, x0_max, x0_n)
x1 = np.linspace(x1_min, x1_max, x1_n)
f = np.zeros((x1_n, x0_n))
for i0 in range(x0_n):
    for i1 in range(x1_n):
        f[i1, i0] = gauss(x0[i0], x1[i1], mu, sigma)
xx0, xx1 = np.meshgrid(x0, x1)  # グリッド座標の作成

# グラフ描画 ----------
plt.figure(figsize=(7, 3))
# 等高線表示
plt.subplot(1, 2, 1)
cont = plt.contour(xx0, xx1, f, levels=15, colors="black")
plt.xlabel("$x_0$", fontsize=14)
plt.ylabel("$x_1$", fontsize=14)
plt.xlim(x0_min, x0_max)
plt.ylim(x1_min, x1_max)
plt.grid()
# サーフェス表示
ax = plt.subplot(1, 2, 2, projection="3d")
ax.plot_surface(
    xx0, xx1, f,
    rstride=2, cstride=2, alpha=0.3, color="blue", edgecolor="black",
)
ax.set_zticks([0.05, 0.10])
ax.set_xlabel("$x_0$", fontsize=14)
ax.set_ylabel("$x_1$", fontsize=14)
ax.view_init(40, -100)
plt.show()
```

Out	# 実行結果は図 4.37 を参照

プログラムを実行すると、図 4.37 のグラフが得られます。分布の中心は、プログラム中で設定した通り、(1, 0.5) にあることがわかります **(A)**。また $\sigma_{01}=1$ としたので、右上がりに広がる分布となりました **(B)**。

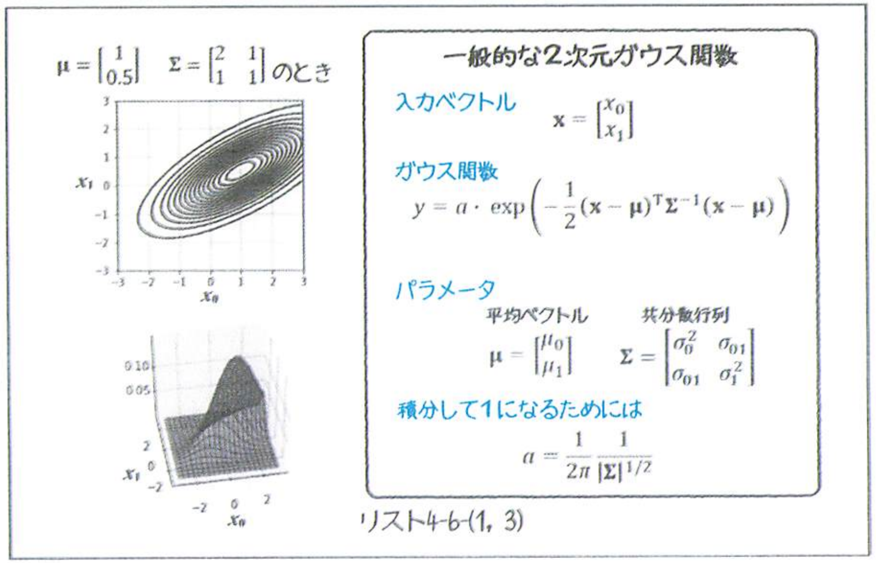


図 4.37：一般的な 2 次元ガウス関数

第 5 章

教師あり学習：回帰

第2問　標準　《共通接線，極大・極小，面積》

$C_1: y=f(x)=x^2-1$

$C_2: y=g(x)=x^3+ax^2+bx+c$　(a, b, c は実数)

(1)　曲線 C_1 の点 $\mathrm{A}(-1,\ -2)$ における接線 ℓ の方程式は，$y-(-2)=f'(-1)\{x-(-1)\}$ と表せるから，$f'(x)=3x^2$，$f'(-1)=3\times(-1)^2=\boxed{3}$ により

$$\ell: y=\boxed{3}x+\boxed{1}\quad すなわち\quad 3x-y+1=0$$

である。また，原点Oと直線 ℓ の距離は，点と直線の距離の公式により

$$\frac{|3\times0-0+1|}{\sqrt{3^2+(-1)^2}}=\frac{1}{\sqrt{10}}=\frac{\sqrt{\boxed{10}}}{10}$$

である。

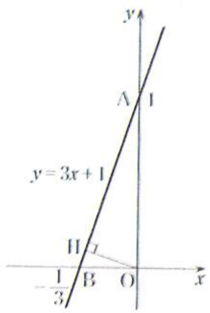
(注)　この距離は図形の性質を利用して求めることもできる。右図の直角三角形OABに対して，三平方の定理から

$$\mathrm{AB}^2=\mathrm{OA}^2+\mathrm{OB}^2=1^2+\left(\frac{1}{3}\right)^2=\frac{10}{9}\qquad \therefore\ \mathrm{AB}=\frac{\sqrt{10}}{3}$$

であり，$\triangle\mathrm{HAO}\backsim\triangle\mathrm{OAB}$ より，$\dfrac{\mathrm{OH}}{\mathrm{OA}}=\dfrac{\mathrm{BO}}{\mathrm{BA}}$ が成り立つから

$$\mathrm{OH}=\frac{\mathrm{OA}\times\mathrm{BO}}{\mathrm{BA}}=1\times\frac{1}{3}\times\frac{3}{\sqrt{10}}=\frac{1}{\sqrt{10}}=\frac{\sqrt{10}}{10}$$

となる。



(2)　曲線 C_2 が点 $\mathrm{A}(-1,\ -2)$ を通ることから

$$g(-1)=-2\quad すなわち\quad -1+a-b+c=-2\quad\cdots\cdots①$$

が成り立つ。また，C_2 の点Aにおける接線（傾きは $g'(-1)$）が直線 ℓ（傾きは3）と一致することと，$g'(x)=3x^2+2ax+b$ より

$$g'(-1)=\boxed{3}\quad すなわち\quad 3-2a+b=3\quad\cdots\cdots②$$

が成り立つ。

②より，$b=\boxed{2}a$ であり，①に代入して，$c=\boxed{a}-\boxed{1}$ となる。

(3)　$a=-2$ のとき，(2)より，$b=2\times(-2)=-4$，$c=(-2)-1=-3$ であるから

$$g(x)=x^3-2x^2-4x-3$$

である。

$$g'(x)=3x^2-4x-4=(3x+2)(x-2)$$

より，$g(x)$ の増減表は右のようになる。

x	$\cdots$	$-\frac{2}{3}$	$\cdots$	2	$\cdots$
$g'(x)$	$+$	0	$-$	0	$+$
$g(x)$	↗	極大 $g\left(-\frac{2}{3}\right)$	↘	極小 $g(2)$	↗

$$g\left(-\frac{2}{3}\right)=-\frac{8}{27}-2\times\frac{4}{9}-4\times\left(-\frac{2}{3}\right)-3$$
$$=\frac{-8-24+72-81}{27}=-\frac{41}{27}$$

$$g(2)=8-8-8-3=-11$$

であるから，関数 $g(x)$ は

$x=\dfrac{\boxed{-2}}{\boxed{3}}$ で極大値 $\dfrac{\boxed{-41}}{\boxed{27}}$ をとり，$x=\boxed{2}$ で極小値 $\boxed{-11}$ をとる。

(4)　(2)より，$b=2a$，$c=a-1$ であるから

$$g(x)=x^3+ax^2+2ax+a-1$$

と表せる。このとき

$$\begin{aligned}f(x)-g(x)&=(x^2-1)-(x^3+ax^2+2ax+a-1)\\&=-ax^2-2ax-a\\&=-a(x^2+2x+1)\\&=-a(x+1)^2\end{aligned}$$

であるが，いま，$a<0$ とするから，$-a(x+1)^2\geqq0$ である。等号は $x=-1$ に対してのみ成り立つ。よって，$-2\leqq x\leqq-1$ において，曲線 C_1 と C_2 および直線 $x=-2$ で囲まれた図形の面積 S_1 は，$S_1=\displaystyle\int_{-2}^{-1}\{f(x)-g(x)\}dx$ と表せる。また，$-1\leqq x\leqq1$ において，曲線 C_1 と C_2 および直線 $x=1$ で囲まれた図形の面積 S_2 は，$S_2=\displaystyle\int_{-1}^{1}\{f(x)-g(x)\}dx$ と表せる。このとき，$S=S_1+S_2$ とおくと

$$S=\int_{-2}^{-1}\{f(x)-g(x)\}dx+\int_{-1}^{1}\{f(x)-g(x)\}dx=\int_{-2}^{1}\{f(x)-g(x)\}dx$$

となるから，$\boxed{ニ}$ に当てはまるものは $\boxed{③}$ である。

これを計算すると

$$S=\int_{-2}^{1}\{f(x)-g(x)\}dx=\int_{-2}^{1}\{-a(x^2+2x+1)\}dx$$
$$=-a\left[\frac{x^3}{3}+x^2+x\right]_{-2}^{1}=-a\left\{\frac{1-(-8)}{3}+(-3)+3\right\}$$
$$=\boxed{-3}a$$

となる。

解説

(1) 接線の方程式，点と直線の距離の公式，いずれも重要かつ頻出である。

> **ポイント　接線の方程式**
>
> 曲線 $y=f(x)$ 上の点 $(t,\ f(t))$ におけるこの曲線の接線の方程式は
>
> $$y-f(t)=f'(t)(x-t)$$

> **ポイント　点と直線の距離の公式**
>
> 点 $\mathrm{P}(x_0,\ y_0)$ と直線 $ax+by+c=0$ の距離 d は
>
> $$d=\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}$$

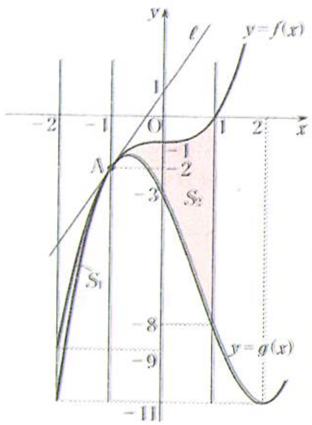
(2) $g(x)=x^3+ax^2+bx+c$ の係数には，a，b，c の3つの未知数があるが，それに対し，条件は

- $y=g(x)$ は点 $\mathrm{A}(-1,\ -2)$ を通る。
- $y=g(x)$ の点Aにおける接線の傾きは3である。

の2つしかないので，a，b，c は決定されない。しかし，b，c それぞれを a で表すことはできる。

(3) $a=-2$ とするから，$g(x)$ の係数がすべて決まる。手順に従って，極大・極小を調べればよい。分数計算をミスしないよう気を付けよう。

(4) $y=f(x)$ と $y=g(x)$ のグラフを描いて，面積を求める図形を確認するのが基本であるが，$g(x)$ には文字係数 a が含まれている（$a=-2$ は(3)だけのこと，ここでは $a<0$ のみが条件）ので，$y=g(x)$ のグラフは描きにくい。$-2\leqq x\leqq -1$，$-1\leqq x\leqq 1$ のそれぞれの区間で，$y=f(x)$ と $y=g(x)$ にはさまれた部分の面積を求めるのであるから，$y=f(x)$ と $y=g(x)$ でどちらが上にあるかが問題である。そこで，〔解答〕のように，$f(x)-g(x)$ を計算してみることになる。これで必要な情報が得られる。ちなみに，$a=-2$ として $y=g(x)$ のグラフを描くと，右図のようになり，$a<0$ の場合の一般的な形が想像できるであろう。



第3問　標準　《漸化式，群数列》

$a_1=1$

$a_{2n}=a_n\quad (n=1,\ 2,\ 3,\ \cdots)$ ……①

$a_{2n+1}=a_n+a_{n+1}\quad (n=1,\ 2,\ 3,\ \cdots)$ ……②

(1) $a_2=1$，$a_3=2$ である。

$n=2$ とすると，①より $a_4=a_2=\boxed{1}$，②より $a_5=a_2+a_3=1+2=\boxed{3}$，

$n=3$ とすると，①より $a_6=a_3=\boxed{2}$，②より $a_7=a_3+a_4=2+1=\boxed{3}$ である。

同様に，$a_{18}=a_9=a_4+a_5=1+3=\boxed{4}$ であり，

$a_{38}=a_{19}=a_9+a_{10}=4+a_5=4+3=\boxed{7}$ である。

(2) 自然数 k に対して，①を連続的に用いると

$$a_{3\cdot 2^k}=a_{3\cdot 2^{k-1}}=a_{3\cdot 2^{k-2}}=\cdots=a_{3\cdot 2}=a_3=2$$

となるから，$\{a_n\}$ の第 $3\cdot 2^k$ 項は $\boxed{2}$ である。

(3) 数列 $\{a_n\}$ の第3項以降を，第 k 群は 2^k 個の項からなるものとして

$$\underset{\text{第1群}}{a_3,\ a_4}\ \Big|\ \underset{\text{第2群}}{a_5,\ a_6,\ a_7,\ a_8}\ \Big|\ \underset{\text{第3群}}{a_9,\ \cdots,\ a_{16}}\ \Big|\ a_{17},\ \cdots$$

のように群に分けるとき，第1群から第 $(k-1)$ 群（k は2以上の自然数）までの項数の和は

$$2^1+2^2+\cdots+2^{k-1}\quad \text{すなわち}\quad \sum_{j=1}^{k-1}2^j=\frac{2(2^{k-1}-1)}{2-1}=\boxed{2}^k-\boxed{2}$$

である。よって，第 k 群の最初の項の直前までに $\{a_n\}$ の項は，$2+(2^k-2)=2^k$ 個あることになるから，第 k 群の最初の項は，$\{a_n\}$ の第 $(2^k+\boxed{1})$ 項である。

また，第 k 群の最後の項は，$\{a_n\}$ の $2^k+2^k=2\times 2^k=2^{k+1}$ 番目の項，すなわち第 2^{k+1} 項である（これらは，$k=1$ に対しても成り立つ）。

$\boxed{ケ}$，$\boxed{シ}$ に当てはまるものは，順に $\boxed{②}$，$\boxed{③}$ である。

第 k 群に含まれるすべての項の和を S_k，第 k 群に含まれるすべての奇数番目の項の和を T_k，第 k 群に含まれるすべての偶数番目の項の和を U_k とするとき

$S_1=a_3+a_4=2+1=\boxed{3}$

$S_2=a_5+a_6+a_7+a_8=3+2+3+a_4=8+1=\boxed{9}$　（①より，$a_8=a_4$）

$T_2=a_5+a_7=3+3=\boxed{6}$

$U_2=S_2-T_2=9-6=\boxed{3}$

である。