

『AI・データサイエンスのための数学入門』学習補助資料

画像を数値として読み、変換する

画素・明るさ・フィルタ・境界・座標変換

1 小さな画像を、行列と見比べる

グレースケール画像は、1 画素に 1 つの明るさを記録したものです。「モノクロ」はここでは白黒 2 値だけでなく、中間の灰色も含む意味で使います。次の例は縦 3、横 4 画素なので、 3×4 行列です。

$$X = \begin{pmatrix} 0 & 64 & 128 & 255 \\ 32 & 96 & 160 & 224 \\ 64 & 128 & 192 & 255 \end{pmatrix}.$$

		1	2	column j	3	4
1	row i	0	64	128	255	
2	row i	32	96	160	224	
3	row i	64	128	192	255	

図 1：各マスの数が画素値です。0 は黒、255 は白。行番号 i は上から下、列番号 j は左から右へ増えます。例えば $x_{2,3} = 160$ です。

カラーでは、同じ位置に 3 つの数がある

1 画素が $(R, G, B) = (255, 0, 0)$ なら赤、 $(0, 255, 0)$ なら緑、 $(0, 0, 255)$ なら青です。3 成分が等しい $(128, 128, 128)$ は灰色になります。

縦 H 、横 W の RGB 画像は、 $H \times W \times 3$ 個の成分をもちます。各成分が 8 ビットなら、非圧縮の画素データ量は $3HW$ バイトです。ファイル全体の容量は、圧縮や付加情報によって異なります。

画素値の行列と、変換を表す行列は役割が違う

X は画像そのもののデータです。一方、座標変換の 2×2 行列 A は、位置を移す規則です。 AX という通常の行列積が、そのまま画像の回転を意味するわけではありません。

線形性を考える際は、画素値を実数まで広げて $\mathbb{R}^{H \times W}$ 上の演算として扱います。0～255 の整数だけの集合は、加法や実数倍について閉じたベクトル空間ではありません。

読み方：2～5 節で演算と図、6 節で本文の 3 問、7～8 節で追加問題と解答を扱います。

2 明るさの演算と、線形性の境界

画素値 x に対する基本操作を、変換後の値 y として整理します。 α, c は画像によらず固定した値とします。

操作	式	クリッピング前の性質
一定値を加える	$y = x + c$	$c \neq 0$ ならアフィンで非線形
一定倍する	$y = \alpha x$	線形。差も α 倍
階調を反転	$y = 255 - x$	アフィンで非線形

画素間の差を見ると、効果がわかる

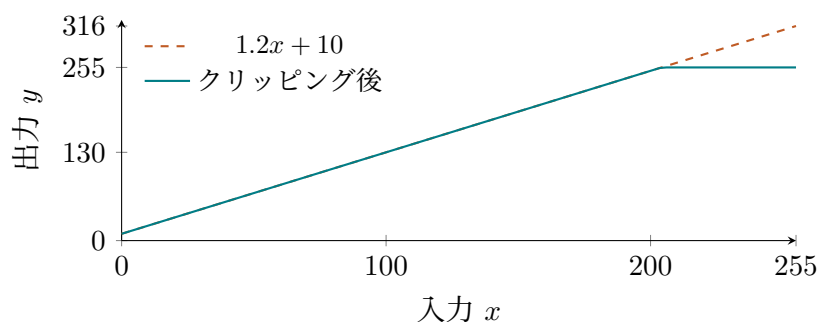
2 つの画素 x_1, x_2 に同じ定数を加えても、差は $(x_2 + c) - (x_1 + c) = x_2 - x_1$ のままです。一方、 $\alpha > 1$ を掛けると差も大きくなります。これが単純なコントラスト調整の見方です。 αx は、平均的な明るさも変化させます。

クリッピングは、四捨五入とは別の処理

範囲外の値を端に収める処理をクリッピングといいます。

$$\text{clip}(y) = \min(255, \max(0, y)).$$

例えば $1.2x + 10$ は、 $x = 100$ で 130、 $x = 240$ で 298 です。後者を 8 ビット範囲に収めると 255 になります。整数への四捨五入などの量子化は、さらに別の操作です。



範囲制限を加えると、線形性は一般に失われる

$L(x) = 2x$ は線形ですが、 $T(x) = \text{clip}(2x)$ は線形ではありません。 $T(100) = 200$ に対して $T(200) = 255 \neq 2T(100)$ です。処理全体の線形性は、飽和や量子化も含めて考えます。

中心値 m の周りのコントラスト調整は $y = m + \alpha(x - m)$ とも書けます。 m を固定したとき、これは一般には定数項 $(1 - \alpha)m$ をもつアフィン写像です。

3 フィルタは、対応する成分の積の和

3×3 の近傍画素を $P = (p_{ij})$ 、カーネルを $K = (k_{ij})$ とすると、中心の出力は

$$y = \sum_{i=1}^3 \sum_{j=1}^3 k_{ij} p_{ij}.$$

これは通常の行列積 KP ではありません。同じ位置の成分を掛け、9 個を足して 1 つの数を作る操作です。

平均フィルタを 1 画素ずつ計算する

例えば $P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 90 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ なら、 $K = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ による中央の出力は $90/9 = 10$ 。局所的に大きな値を、周囲の値と平均して弱めます。

線形性の証明

固定カーネルによる処理を L_K と書きます。同じ位置の 2 つの近傍 P, Q に対して

$$\sum_{i,j} k_{ij} (ap_{ij} + bq_{ij}) = a \sum_{i,j} k_{ij} p_{ij} + b \sum_{i,j} k_{ij} q_{ij}.$$

したがって、同じ境界処理のもとで画像全体についても $L_K(aX + bY) = aL_K(X) + bL_K(Y)$ となります。ただし、ここではクリッピング・量子化を含めず、境界での拡張も線形とします。

畳み込みと相互相関の違い

カーネルをそのまま重ねる上の操作は、厳密には**相互相関**です。数学的な畳み込みではカーネルを 180 度回転させてから重ねます。深層学習では相互相関も慣習的に「畳み込み」と呼びます。本文のカーネルはいずれも 180 度回転で不変なので、両者の計算結果は一致します。

重みの和から、一定画像への反応がわかる

近傍の全画素が c なら、出力は $c \sum_{i,j} k_{ij}$ です。

- 重みの和が 1：一定の明るさ c を保ちます。平均フィルタが例です。
- 重みの和が 0：一定領域では 0 になります。変化を検出するフィルタに使われます。

本文の W では、各行に $-1, 2, -1$ が並びます。中央列が明るく両隣が暗いと正の大きな応答が出ます。縦の線や左右方向の急な変化に反応しますが、すべての方向の境界に同じように反応するわけではありません。

4 画像全体で、平滑化と輪郭の反応を見る

次の図は、一定の背景、長方形、縦の明るい帯から作った小さな人工画像です。処理の違いが追えるよう、入力と表示範囲を明示します。

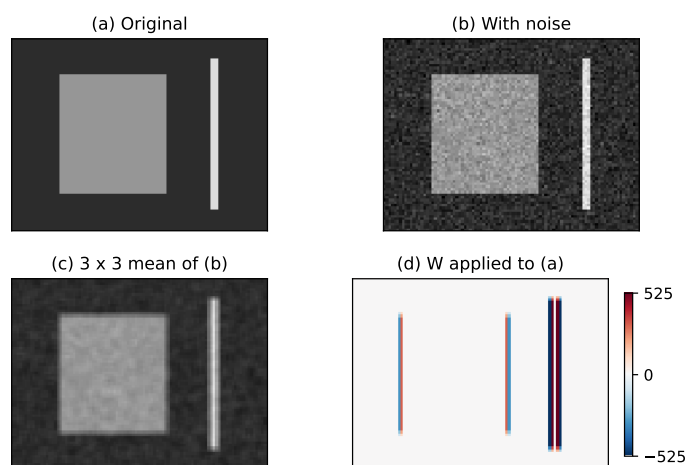


図 2：(a) 元画像。(b) 標準偏差 15 の独立な正規ノイズを加え、0～255 にクリップ。(c) (b) に平均フィルタを適用。(d) 元画像 (a) に本文の W を適用。赤が正、青が負、白が 0。境界は反射で拡張。乱数シード 21。

ノイズは減るが、細部もなまる

平均フィルタは細かな変動を抑える一方で、細い線や境界もぼかします。理想的に 9 画素のノイズが独立で分散 σ^2 なら、その平均のノイズ分散は $\sigma^2/9$ です。ただし、重なり合う近傍から得た出力同士は一般に相関をもちます。

画像の端では、近傍がはみ出す

方法	内容と結果
有効範囲のみ	3×3 、移動幅 1 なら出力は $(H - 2) \times (W - 2)$
ゼロ埋め	外側を 0 とする。平均フィルタで端が暗くなることがある
端の複製・反射	端の値や鏡映した値で補う。一定画像を一定に保てる

符号付きの出力を、そのまま 8 ビット画像にしない

輪郭フィルタの出力は負や 255 超にもなります。計算では符号付き実数で保持し、表示するときに別途スケールを決めます。絶対値化やクリッピングを加えると、その表示処理まで含めた写像は一般に線形ではなくなります。

5 座標を動かすことと、画素値を作ること

画素値変換は「同じ場所の明るさ」を変えます。幾何学的変換は「どこに対応させるか」を変えます。座標 $p = (x, y)^\top$ の移動は

$$p' = Ap + b$$

と書けます。以下の図では、数学の座標として右向きを x の正、上向きを y の正に取ります。

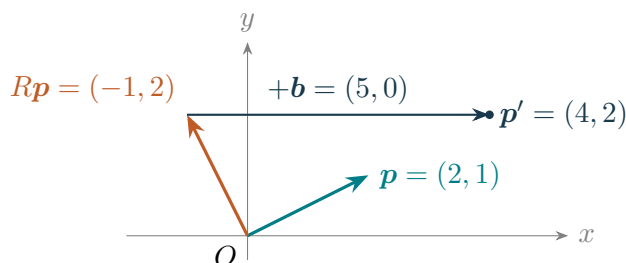


図 3：90 度反時計回り回転の後で平行移動する例。順序を逆にすると、結果は一般に異なります。

画像の行番号は、普通は下向きに増える

画像では $x = j$ 、 $y = i$ として右・下を正に取ることがあります。この座標に数学の回転行列をそのまま使うと、画面上の回転方向が逆に見えます。原点・軸の向き・画素中心の位置を先に決めましょう。

画像の中心 c の周りに回す場合は

$$p' = R(p - c) + c = Rp + (I - R)c$$

です。一般には、原点中心の線形変換に平行移動を加えたアフィン変換になります。

新しい格子点に、どの明るさを入れるか

回転後の位置は整数格子点と一致するとは限りません。可逆な A なら、出力側の各点から $p = A^{-1}(p' - b)$ を求め、元画像の値を補間する方法が使えます。単に入力画素を前方へ移すだけでは、出力に穴ができることがあります。

補間の最小例

1 次元で、位置 0 の値が 20、位置 1 の値が 100 なら、位置 0.25 の線形補間値は $(1 - 0.25)20 + 0.25 \cdot 100 = 40$ 。2 次元の双線形補間では、周囲 4 画素の重み付き和を使います。

変換座標と補間重みを固定すれば、補間は画素値に関して線形にできます。しかし、離散化・切り出し・縮小などで情報が失われるため、座標行列が可逆でも画像を完全復元できるとは限りません。

6 本文の練習問題を解く

問 1：平均フィルタと輪郭フィルタ

$$X = \begin{pmatrix} 10 & 20 & 30 \\ 20 & 40 & 60 \\ 30 & 60 & 90 \end{pmatrix}.$$

平均フィルタは全 9 成分の平均なので、

$$\frac{10 + 20 + 30 + 20 + 40 + 60 + 30 + 60 + 90}{9} = \frac{360}{9} = \boxed{40}.$$

次に $K = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$ は、中央を 4 倍し、上下左右の値を引きます。

$$4 \cdot 40 - (20 + 20 + 60 + 60) = 160 - 160 = \boxed{0}.$$

平均フィルタは近傍をならし、 K は中央と周囲との差に反応します。この例では、中央 40 が上下左右の平均 40 と一致しているため 0 になります。

応答 0 でも、近傍が一定とは限らない

上の 9 画素は同じ値ではありません。それでも、この位置では重み付き和が相殺されて 0 です。「フィルタ値が 0 だから変化がない」とは断定できません。応答はカーネルと局所的なパターンの組合せで決まります。

問 2： $Y = 1.2X + 10J$ の意味

各画素を 1.2 倍してから 10 を加えます。クリッピング前には、画素間の差が 1.2 倍になり、単純な意味でコントラストが強まります。非負の入力に対して出力値は入力より大きくなり、明るさも増します。

$T(0) = 10J \neq 0$ なので線形ではなく、アフィン写像です。255 を超える画素はクリッピングで飽和し、明るい部分の差が失われることがあります。例えば 220 と 240 は変換前には差 20 ですが、変換・クリッピング後はどちらも 255 です。

問 3：90 度回転の後に平行移動

数学の通常の座標系で、回転行列は

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad R \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

その後に $\mathbf{b} = (5, 0)^\top$ を加えるので、

$$\mathbf{p}' = \begin{pmatrix} -1 \\ 2 \end{pmatrix} + \begin{pmatrix} 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}.$$

順序を逆にした $R(\mathbf{p} + \mathbf{b})$ は $R(7, 1)^\top = (-1, 7)^\top$ となり、同じ結果ではありません。

7 追加の練習問題

特に断らない限り、フィルタはカーネルをそのまま重ねる成分積の和とし、クリッピングや整数への丸めは行いません。

問 4：配列の形と画素数

縦 480、横 640 の RGB 画像について、配列の形、画素数、成分数を求めよ。各成分 8 ビットの非圧縮画素データは何バイトか。

問 5：明るさ変換と飽和

$X = \begin{pmatrix} 0 & 100 \\ 200 & 250 \end{pmatrix}$ に $Y = 1.5X + 20J$ を適用せよ。次に 0~255 へクリッピングした結果を求めよ。クリッピング前後で失われる画素間の差を 1 つ示せ。

問 6：一定画像へのフィルタ応答

近傍の全 9 画素が 80 である。平均フィルタと、本文の $W = \begin{pmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{pmatrix}$ の出力をそれぞれ求めよ。

問 7：縦の線への反応

$P = \begin{pmatrix} 0 & 100 & 0 \\ 0 & 100 & 0 \\ 0 & 100 & 0 \end{pmatrix}$ に、平均フィルタと問 6 の W を中央に適用せよ。平均フィルタの結果との違いを説明せよ。

問 8：端での平均値

3×3 の全画素が 90 の画像の左上画素に平均フィルタを適用する。画像外を 0 とする場合と、端の値を複製する場合の値を求めよ。

問 9：相互相関と畳み込み

1 次元の近傍 (10, 20, 40) とカーネル (-1, 0, 1) を考える。そのまま重ねた相互相関と、カーネルを逆順にした畳み込みの出力を求めよ。

問 10：座標変換と逆向きの対応

$p' = Ap + b$ 、 $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ 、 $b = (1, -1)^T$ とする。 $p = (3, 2)^T$ の像を求めよ。また、出力座標 $(5, 2)^T$ に対応する入力座標を求めよ。

発展：問 11

各成分が $[0, 255]$ にある画像に固定カーネルを適用する。カーネルの全重みが非負で和が 1 なら、出力も $[0, 255]$ にあることを示せ。負の重みを許す場合、和が 1 でもこの結論が成り立たない例を挙げよ。

8 追加問題の解答と考え方

問 4

配列の形は $480 \times 640 \times 3$ 。画素数は $480 \cdot 640 = 307200$ 、成分数は 921600。1 成分 1 バイトなので、非圧縮画素データは 921600 バイト です。

問 5

$$Y = \begin{pmatrix} 20 & 170 \\ 320 & 395 \end{pmatrix}, \quad \text{clip}(Y) = \begin{pmatrix} 20 & 170 \\ 255 & 255 \end{pmatrix}.$$

下段の 2 画素の差は、クリッピング前は 75 ですが、後は 0 です。飽和により、もとの違いが区別できなくなります。

問 6

平均フィルタは $9 \cdot 80 / 9 = \text{80}$ 。W の重みの和は 0 なので出力は 0 です。

問 7

平均は $300 / 9 = \text{100/3}$ 。W では中央列の 3 画素にそれぞれ 2 を掛けるので、出力は 600 です。平均は値をならしますが、W は両隣より明るい中央列を強く検出します。600 という値は、計算段階ではそのまま保持します。

問 8

ゼロ埋めでは、近傍 9 か所のうち画像内にあるのは 4 か所です。したがって $4 \cdot 90 / 9 = \text{40}$ 。端の複製では 9 か所とも 90 になるので 90 です。境界処理が結果を変える例です。

問 9

相互相関は $-10 + 40 = \text{30}$ 。畳み込みでは $(1, 0, -1)$ を使うので $10 - 40 = \text{-30}$ 。非対称カーネルでは、反転の有無が重要です。

問 10

$\mathbf{p}' = (6, 2)^\top + (1, -1)^\top = \text{(7, 1)^\top}$ 。逆向きには平行移動を先に取り消します。

$$\mathbf{p} = A^{-1}(\mathbf{p}' - \mathbf{b}) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \text{\begin{pmatrix} 2 \\ 3 \end{pmatrix}}.$$

問 11

各画素 x_i と重み w_i が $0 \leq x_i \leq 255$ 、 $w_i \geq 0$ 、 $\sum_i w_i = 1$ を満たすなら、

$$0 \leq \sum_i w_i x_i \leq \sum_i w_i \cdot 255 = 255.$$

一方、重み $(2, -1)$ は和が 1 ですが、画素 $(255, 0)$ に作用させると 510 になります。重みの和が 1 という条件だけでは、範囲内に収まる保証はありません。