

数理アルゴリズムとシミュレーション 演習課題 8

提出期限: 2020/12/17 23:59

以下の課題を行い、レポートを提出すること。レポートの作成に関しては、manaba の「演習課題」ページ内の項目をしっかりと確認すること。また、レポートの作成にあたって参考にした文献や Web ページはその出典を明示すること。

課題 1

(1-1) \mathbf{x} が最小二乗問題 $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|_2$ の解であるとき、 \mathbf{x} は正規方程式 の解である。空欄に当てはまるものを選べ。

(a) $A^T A \mathbf{x} = A \mathbf{b}$

(b) $A^T A \mathbf{x} = A^T \mathbf{b}$

(c) $AA^T \mathbf{x} = A \mathbf{b}$

(d) $AA^T \mathbf{x} = A^T \mathbf{b}$

(1-2) 行列 $A \in \mathbb{R}^{m \times n}$, $\text{rank } A = n$ の特異値分解は

$$A = U \Sigma V^T, \quad U \in \mathbb{R}^{m \times n}, \quad \Sigma \in \mathbb{R}^{n \times n}, \quad V \in \mathbb{R}^{n \times n}, \\ U^T U = I, \quad V V^T = V^T V = I$$

と表せる。このとき、最小二乗問題 $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|_2$ の解として正しいものを選べ。

(a) $\mathbf{x} = V \Sigma U^T \mathbf{b}$

(b) $\mathbf{x} = U \Sigma V^T \mathbf{b}$

(c) $\mathbf{x} = V \Sigma^{-1} U^T \mathbf{b}$

(d) $\mathbf{x} = U \Sigma^{-1} V^T \mathbf{b}$

(1-3) (1-2) と同様に行列 A の特異値分解が与えられたとする。また、 $B \in \mathbb{R}^{m \times l}$ とする。このとき、最小二乗問題 $\min_{X \in \mathbb{R}^{n \times l}} \|B - AX\|_F$ の解として正しいものを選べ。

(a) $X = V \Sigma U^T B$

(b) $X = U \Sigma V^T B$

(c) $X = V \Sigma^{-1} U^T B$

(d) $X = U \Sigma^{-1} V^T B$

課題 2

手書き数字の画像データセットである MNIST に対して、線形回帰により画像認識を行う。MNIST のデータファイルは manaba にアップロードされているため、ダウンロード

して用いること。ファイルを読み込むと二次元配列 $X, Y, \text{test_X}, \text{test_Y}$ が得られる。変数 X は訓練画像データであり、各列に 1 枚の訓練画像を格納している。変数 Y は各訓練画像が属するグループを示している。変数 test_X はテスト画像データであり、各列に 1 枚のテスト画像を格納している。変数 test_Y は各テスト画像が属するグループを示している。今回の課題では訓練データとテストデータをそれぞれ 10,000 個ずつ用いる。

MATLAB 上で `imshow(reshape(X(:, i), [28, 28]))` と実行すると行列 X の先頭から i 番目の列に格納されている訓練データを画像として表示することができる。

(2-1) 行列 $X \in \mathbb{R}^{m \times n}, Y \in \mathbb{R}^{10 \times n}$ に対し、最小二乗問題

$$\min_{W \in \mathbb{R}^{10 \times m}} \|Y - W\tilde{X}_k\|_F$$

を特異値分解を用いて解く MATLAB プログラムを作成せよ。ここで、 \tilde{X}_k は、 X の最大から k 個の特異値と対応する特異ベクトルで作られる低ランク近似行列であるとする。また、 k は行列 X の特異値を $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$ としたとき、 $\sigma_k/\sigma_1 \geq 10^{-14}$ を満たすような最大の整数とする。ここでは $m = 784, n = 10,000$ である。

(2-2) 得られた行列 W を用いてテストデータの画像がどのグループに属するか判定することができる。manaba にアップロードされている `check.m` を用いて (2-1) で得られた W に対して判定処理を行い、画像認識の正答率を求めよ。`check(W, test_X, test_Y)` と実行すると正答率を求めることができる。

(2-3) (2-2) では訓練データを 10,000 個用いて画像認識を行ったが、訓練データ数を変えることで画像認識の正答率は変化する。訓練データ数を行列の先頭の列から 1,000, 2,000, ..., 10,000 個としたときの画像認識の正答率を求める MATLAB プログラムを作成し、それぞれの訓練データ数における正答率をグラフに描画せよ。

(2-4) 行列 X の低ランク近似を行う際の整数 k を、 $\sigma_k/\sigma_1 \geq \delta$ を満たすような最大の整数とする。このとき、しきい値 δ を 0.005, 0.01, 0.015, ..., 0.1 と変えたときの画像認識の正答率を求める MATLAB プログラムを作成し、各しきい値における正答率をグラフに描画せよ。ただし、訓練データは 10,000 個用いること。