

## 課題2

---

(2-1)

次のようなコードを作成した。

```
-----  
load("data_MNIST.mat");  
m = 784;  
n = 10000;  
  
%%%%%%%%%%%% X_kを求める  
tol = 1e-14;  
[U,S,V] = svd(X);  
s = svd(X);  
s_max = s(1);  
  
for i=1:m  
    val = s(i)/s_max;  
    if val < tol  
        i = i-1;  
        break;  
    end  
end  
  
k = i;  
  
U_k = zeros(m, k);  
for j=1:k  
    U_k(:,j)=U(:,j);  
end  
  
V_k = zeros(n, k);  
for j=1:k  
    V_k(:,j)=V(:,j);  
end  
  
sk = zeros(1,k);  
for j=1:k  
    sk(j)=s(j);  
end  
  
sig_k = diag(sk);  
  
tmp = sig_k\transpose(U_k);
```

```
X_k = V_k *tmp;
```

```
%%%%%%%%%%%% Wを求める
```

```
W = Y*X_k;
```

```
-----
```

---

(2-2)

(2-1)で得られたWを用いて次のように実行した。

```
-----
```

```
%%%%%%%%%%%% Wをテストする
```

```
check(W, test_X, test_Y)
```

```
-----
```

結果は次のようになった。

```
-----
```

```
ans =
```

```
0.8305
```

```
-----
```

正答率が83.05%であることを示している。

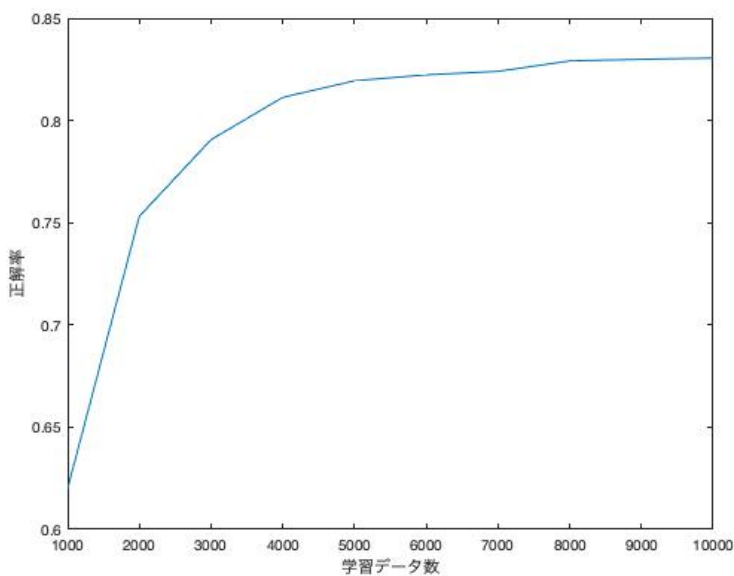
(2-3)

次のようなコードを作成した。

```
-----  
load("data_MNIST.mat");  
m = 784;  
n = 1000;  
  
for w=1:10  
    n_tmp = w*1000;  
    xval(w)=n_tmp;  
    %X_tmp  
    X_tmp = zeros(m,n_tmp);  
    for j=1:n_tmp  
        X_tmp(:,j)=X(:,j);  
    end  
    %Y_tmp  
    Y_tmp = zeros(10,n_tmp);  
    for j=1:n_tmp  
        Y_tmp(:,j)=Y(:,j);  
    end  
  
    %%%%%%%%% X_tmp_kを求める  
    tol = 1e-14;  
    [U,S,V] = svd(X_tmp);  
    s = svd(X_tmp);  
    s_max = s(1);  
  
    for i=1:m  
        val = s(i)/s_max;  
        if val < tol  
            i = i-1;  
            break;  
        end  
    end  
  
    k = i;  
    U_k = zeros(m,k);  
    for j=1:k  
        U_k(:,j)=U(:,j);  
    end  
  
    V_k = zeros(n_tmp, k);  
    for j=1:k
```

```
V_k(:,j)=V(:,j);  
end  
  
sk = zeros(1,k);  
for j=1:k  
    sk(j)=s(j);  
end  
  
sig_k = diag(sk);  
  
tmp = sig_k\transpose(U_k);  
  
X_tmp_k = V_k *tmp;  
  
%%%%%%%%%%%% Wを求める  
  
W_tmp = Y_tmp*X_tmp_k;  
  
%%%%%%%%%%%% Wをテストする  
ch(w)=check(W_tmp, test_X, test_Y);  
end  
ch  
plot(xval,ch)  
xlabel('学習データ数')  
ylabel('正解率')  
-----
```

これを実行すると、図1が描画された。



訓練データ数が多くなると正答率が上がることがわかる。

図1 訓練データ数における正答率

(2-4)

次のようなコードを作成した。

```
-----  
load("data_MNIST.mat");  
m = 784;  
n = 10000;  
tol_0 = 0.005;  
  
for w=1:20  
    tol = tol_0*w;  
    xval(w)=tol;  
    %%%%%%%%% X_kを求める  
    [U,S,V] = svd(X);  
    s = svd(X);  
    s_max = s(1);  
  
    for i = 1:m  
        val = s(i)/s_max;  
        if val < tol  
            i = i-1;  
            break;  
        end  
    end  
  
    k = i;  
  
    U_k = zeros(m, k);  
    for j=1:k  
        U_k(:,j)=U(:,j);  
    end  
  
    V_k = zeros(n, k);  
    for j=1:k  
        V_k(:,j)=V(:,j);  
    end  
  
    sk = zeros(1,k);  
    for j=1:k  
        sk(j)=s(j);  
    end  
  
    sig_k = diag(sk);
```

```
tmp = sig_k\transpose(U_k);

X_k = V_k *tmp;

%%%%%%%%%%%% Wを求める

W = Y*X_k;

%%%%%%%%%%%% Wをテストする
ch(w)=check(W, test_X, test_Y);
end
ch
plot(xval,ch)
xlabel('しきい値')
ylabel('正解率')
-----
```

これを実行すると図2が得られた。

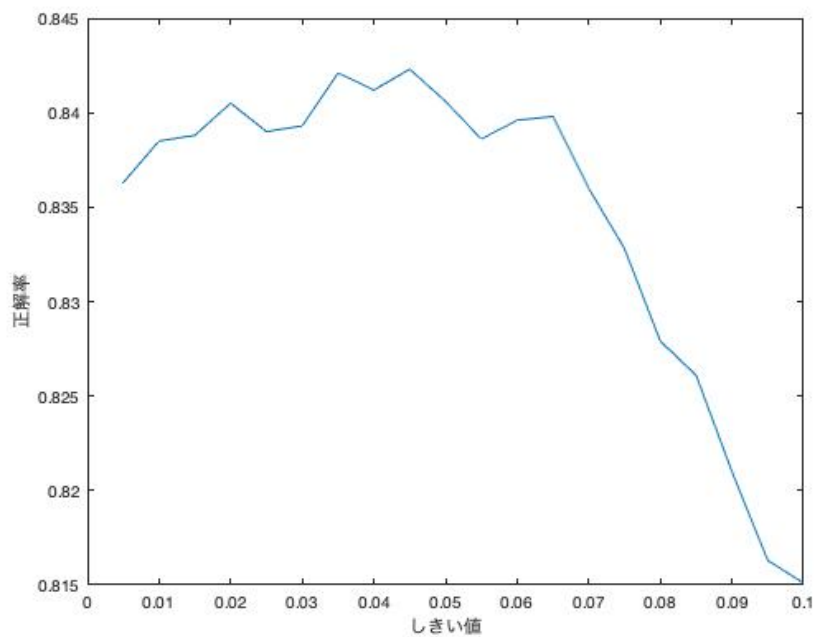


図2 各しきい値における正答率

しきい値が0.065を超えたあたりから正答率が下がっていくことがわかる。