

知的システム論第 10 回レポート

37186305

航空宇宙工学専攻修士一年

荒居秀尚

2018 年 12 月 13 日

1 宿題

Julia 1.0.0 により MNIST dataset のうち test データを次元削減・クラスタリング・可視化した。実装は以下の通りである。

```
1 using LinearAlgebra
2 using Flux.Data.MNIST
3 using Plots
4
5 function PCA(X, m)
6     d = size(X)[1]
7     n = size(X)[2]
8     C = zeros((d, d))
9     for i in 1:n
10         C += X[:, i] * X[:, i]'
11     end
12
13     eigs = eigen(C)
14     T = zeros((m, d))
15     last = d - m + 1
16     for (i, j) in enumerate(d:-1:last)
17         vec = eigs.vectors[:, j]
18         T[i, :] = vec'
19     end
20     T * X
21 end
22
23 function converged(array, tol)
24     mean = 0.0
25     for i in 1:length(array)
26         mean += sum(array[i].^2)
27     end
28     mean < tol
29 end
```

```

30
31
32 function calc_center(X)
33     n = size(X)[2]
34     ret = [0, 0]
35     for i in 1:n
36         ret += X[:, i]
37     end
38     ret ./ n
39 end
40
41
42 function KMeans(X, c, tol=1e-5)
43     d = size(X)[1]
44     n = size(X)[2]
45     centers = [rand(2,1) for _ in 1:c]
46     before = [[100, 100] for _ in 1:c]
47     cpreds = zeros(n)
48     while !converged(before - centers, tol)
49         for i in 1:n
50             norms = [sum((X[:, i] - center).^2) for center in centers]
51             label = argmin(norms)
52             cpreds[i] = label
53         end
54         before = centers
55         idxs = [findall(cpreds .== i) for i in 1:c]
56         centers = [ifelse(X[idx] != [], calc_center(X[:, idx]), before[i])
                    for (i, idx) in enumerate(idxs)]
57     end
58     return centers, cpreds
59 end
60
61 imgs = MNIST.images(:test)
62 X = hcat(float.(vec.(imgs))...)
63 X_pca = PCA(X, 2)
64 centers, cpreds = KMeans(X_pca, 10)
65 p = plot(xlabel="x", ylabel="y", title="Compressed image of MNIST")
66 colors = [:green, :blue, :red, :yellow, :pink, :black, :gold, :silver, :brown
67         , :purple]
68 for (i, c) in enumerate(colors)
69     idx = findall(cpreds .== i)
70     scatter!(X_pca[1, idx], X_pca[2, idx], color=c, label=string(i))
71 end
72 png("scatter")
73
74 labels = MNIST.labels(:test)
75 p = plot(xlabel="x", ylabel="y", title="Compressed image of MNIST with true
76         label")
77 colors = [:green, :blue, :red, :yellow, :pink, :black, :gold, :silver, :brown

```

```

, :purple]
76 for (i, c) in enumerate(colors)
77     idx = findall(labels .== i)
78     scatter!(X_pca[1, idx], X_pca[2, idx], color=c, label=string(i))
79 end
80 png("scatter_true")

```

この結果が以下ようになった。

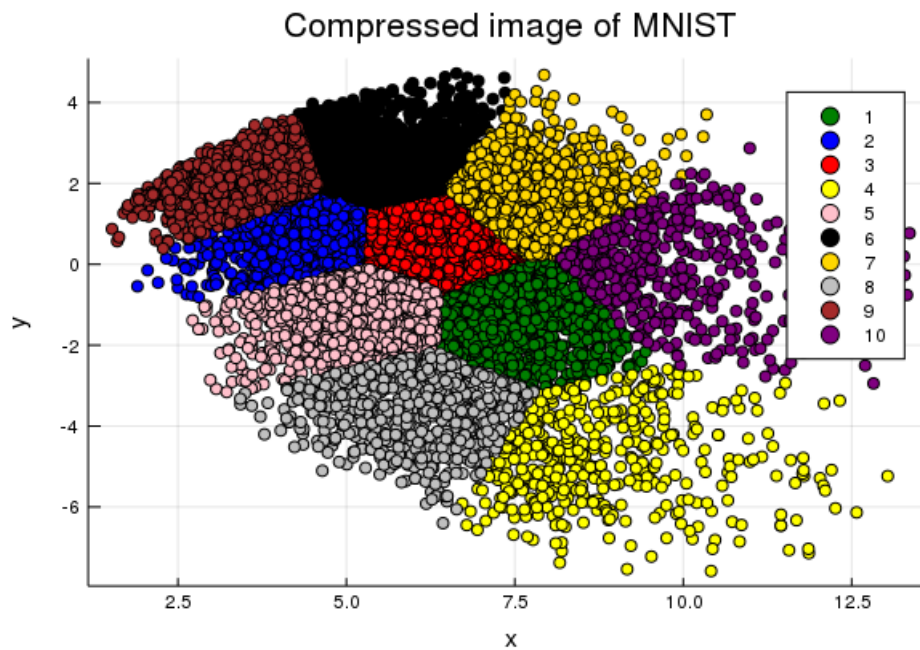


図 1.1: KMeans によるクラスタリング

この二枚の図を見比べるとそもそも 784 次元空間のものを 2 次元に射影して考えることに無理があるとわかる。また、KMeans の境界付近が直線になっており、ボロノイ境界が見えることがわかる。

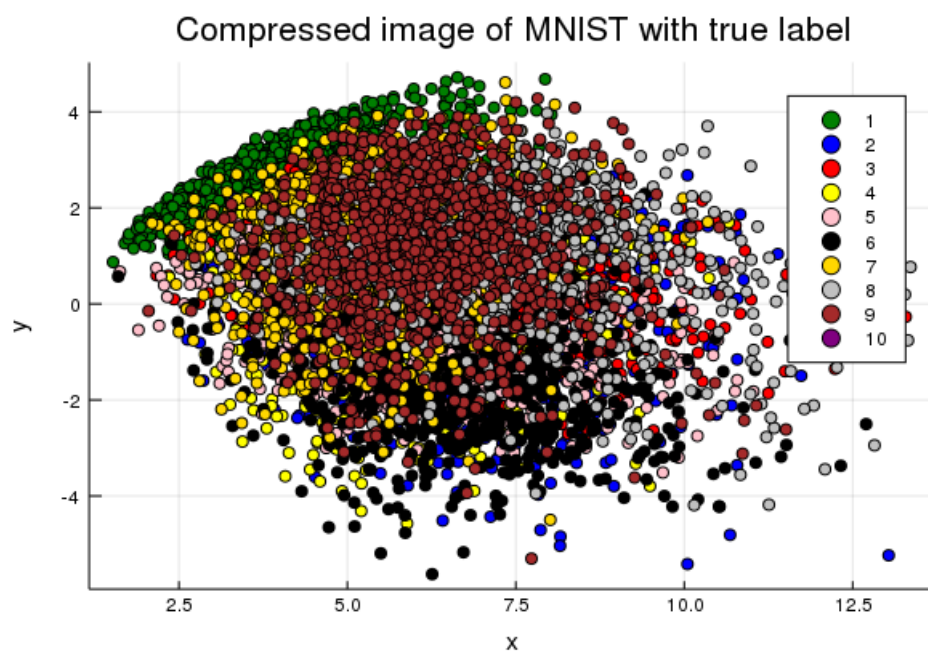


図 1.2: 正解ラベルごとに色分けした図