

lokaverk

April 5, 2024

```
[1]: import random
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
[2]: #Hledur fashion mnist gagnasettid
#hafid path sem toma strenginn ef skrain er i somu moppu og tid erud ad vinna i
def load_f_mnist(path, kind="train"):
    import os
    import gzip

    """Load MNIST data from `path`"""
    labels_path = os.path.join(path, "%s-labels-idx1-ubyte.gz" % kind)
    images_path = os.path.join(path, "%s-images-idx3-ubyte.gz" % kind)

    with gzip.open(labels_path, "rb") as lbpath:
        labels = np.frombuffer(lbpath.read(), dtype=np.uint8, offset=8)

    with gzip.open(images_path, "rb") as imgpath:
        images = np.frombuffer(imgpath.read(), dtype=np.uint8, offset=16).
        ↪reshape(len(labels), 784)

    return images, labels

#Hledur mnist gagnasettid
#hafid path sem toma strenginn ef skrain er i somu moppu og tid erud ad vinna i
def load_mnist(path, kind="train"):
    import os
    import gzip

    """Load MNIST data from `path`"""
    labels_path = os.path.join(path, "mnist_%s-labels-idx1-ubyte.gz" % kind)
    images_path = os.path.join(path, "mnist_%s-images-idx3-ubyte.gz" % kind)

    with gzip.open(labels_path, "rb") as lbpath:
        labels = np.frombuffer(lbpath.read(), dtype=np.uint8, offset=8)
```

```

with gzip.open(images_path, "rb") as imgpath:
    images = np.frombuffer(imgpath.read(), dtype=np.uint8, offset=16).
    ↪reshape(len(labels), 784)

return images, labels

```

```

[4]: X, labels = load_mnist("lokaverk") #mnist
X_f, labels_f = load_f_mnist("lokaverk") #fashion mnsist

```

```

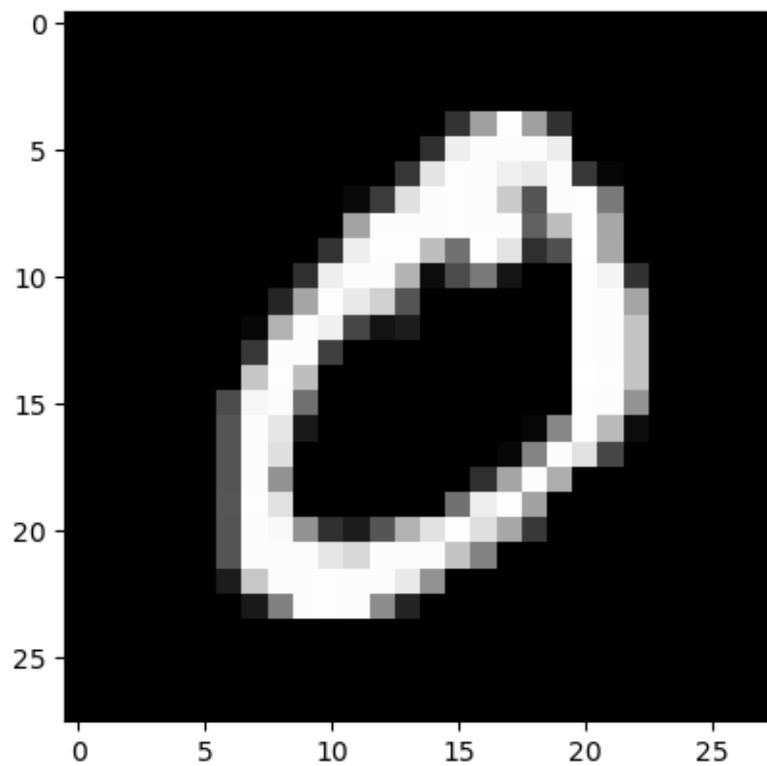
[5]: def plot_one(X,i):
    one = X[i].reshape(28, 28)
    fig = plt.figure
    plt.imshow(one, cmap="gray")
    plt.show()

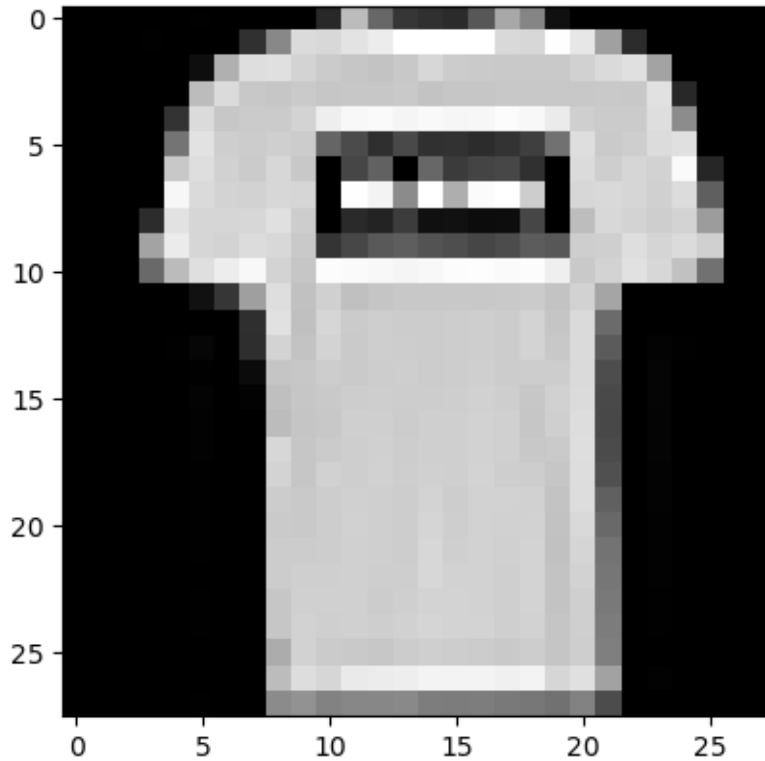
```

```

[6]: plot_one(X,10)
plot_one(X_f,101)

```





```
[8]: def run_kmeans(data, k, max_iter=100):
```

```
def assign(data, centroids):
```

```
def update(data,assignments, centroids):
```

```
[12]: # her erum vid ad keyra kmeans a fashion mnist gagnasettid  
#assignments = run_kmeans(X_f,10)  
assignments_orig = run_kmeans(X,10)
```

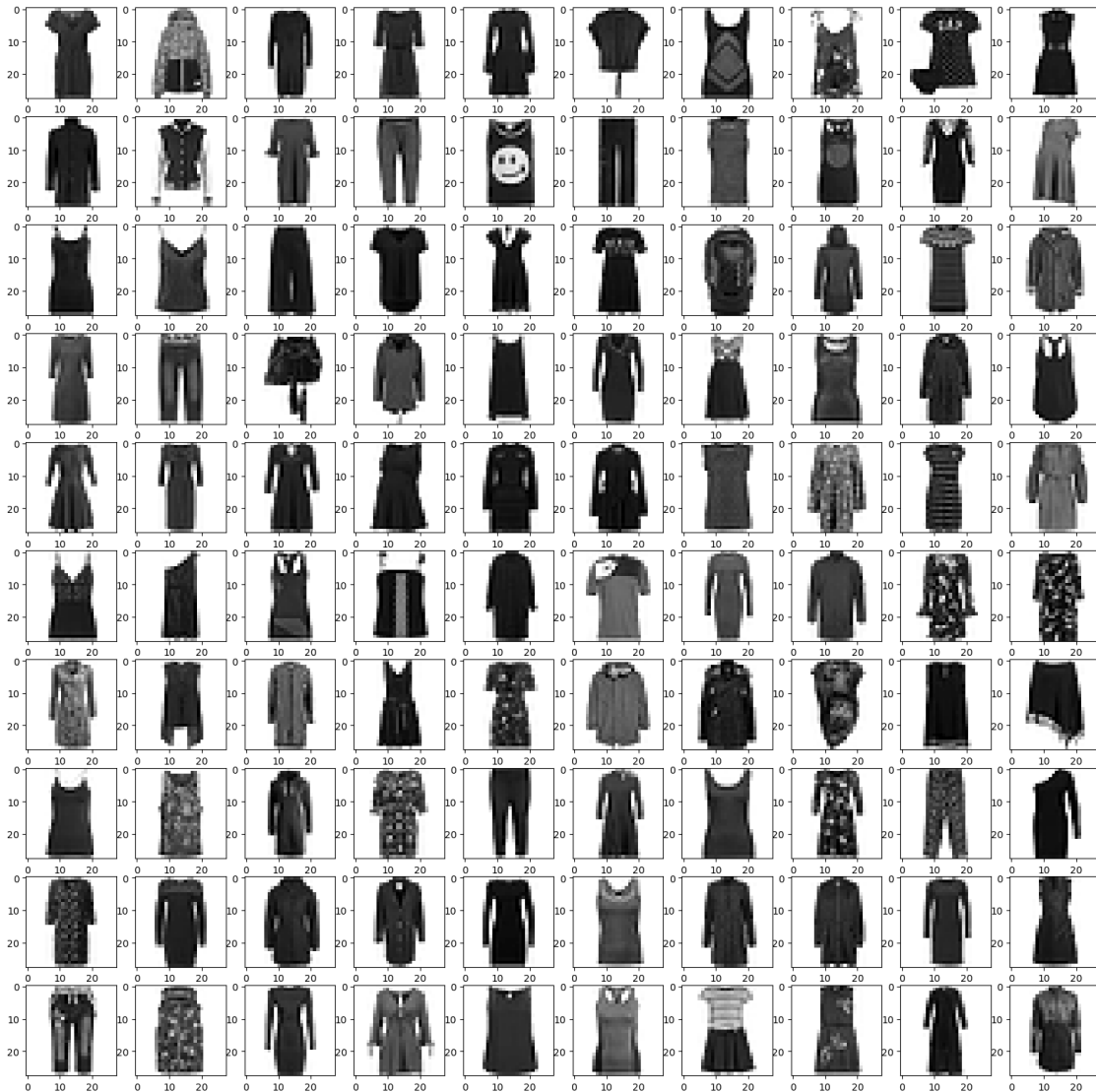
```
1  
2  
3  
4  
5  
6  
7  
8  
9
```

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

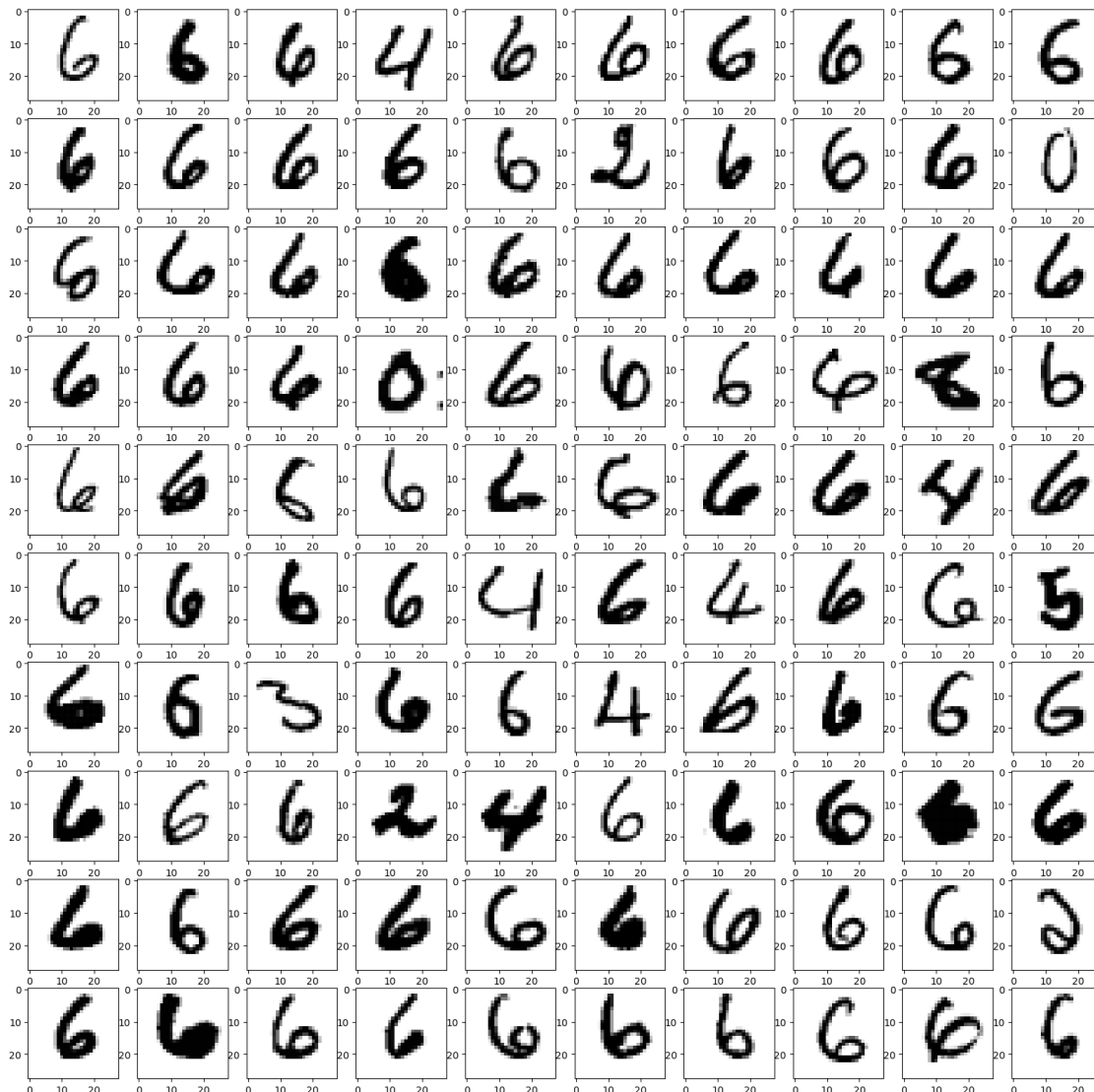
```
[7]: def plot_result(X, clust, output):  
      plt.figure(figsize=(20, 20))  
      for i in range(1, 101):  
          plt.subplot(10, 10, i)  
          plt.imshow(X[np.where(output == clust)][i].reshape(28, 28), cmap=plt.cm.  
→ binary)
```

```
plt.show()
```

```
[11]: #her eru fyrstu 100 stokin sem við reikniritið mitt flokkaði i flokk 2.  
# sjaum að þetta er líklegast kjóla flokkurinn en eitthvað af bolum og buxum  
      ↪ hafa dottið með. Smá óhjákvæmilegt.  
plot_result(X_f,2,assignments)
```



```
[14]: #her eru fyrstu 100 stokin sem við reikniritið mitt flokkaði i flokk 3 i  
      ↪ upprunalega mnist (tölurnar þas).  
# sjaum að þetta er talan 6. Hinsvegar hafa laumast inn einn tvistur og nokkrir  
      ↪ fjargar og slíkt. Samt nokkuð gott.  
plot_result(X,3,assignments_orig)
```



Flokkar stemma ekki endilega milli labela, sem er rétta svarið og svo assignments hjá okkur. Þeas við gætum sett alla tvista í flokk 2 en í labels(rétta svarinu) er það í flokk 5. Við þurfum því aðeins að umrita. Hvernig getum við synt hversu nákvæmt líkanið okkar er. Úthlutum hverjum flokk í giskinu okkar, assignments hér að ofan

```
[16]: #ef assignments eru lokagiskid okkar og labels eru retta svarid sem vid faum
      ↪tegar vid tokum inn gognin
      #og cat er einhver flokkur
      #tetta fall tetta fall okkur besta gisk a hvada flokkur tetta er i labels.
      def get_real_category(assignment,labels,cat):
          a = labels[np.where(assignment==cat)]
          unique, counts = np.unique(a, return_counts=True)
          return dict(zip(unique, counts))
```

```
[20]: #get_real_category(assignments_orig, labels, 1)
d = get_real_category(assignments_orig, labels, 1)
print(d)
```

```
{0: 23, 1: 53, 2: 165, 3: 337, 4: 28, 5: 572, 6: 10, 7: 50, 8: 3374, 9: 52}
```

```
[27]: #Sjaum að ofan að okkar flokkur 1 er í raunveruleikanum flokkur 8.
#sem tyðir að við flokkum í flokk 1 með um 72% nákvæmni
d[8]/sum(d.values())
```

```
[27]: 0.7234133790737565
```

Við flokkum í flokk 1 með um 72% nákvæmni. Sem ef má sjá að neðan stemmir því þetta er “raunverulega” flokkur 8. Það fylgja einhverjir, ásar, fimmur og þristar með.

```
[28]: plot_result(X, 1, assignments_orig)
```

