# Machine Learning Algorithms: From Math to Code
# Assignment for Feature Selection

Kang Wang, Yuheng Zhang

## 1 Problem Set 1

For this problem set, you should complete the missing code for Ch9.m and Ch10.m, and submit them with a report describing your results in a compressed .zip file on canvas.

### 1.1 Ch9.m

After you finish Ch9.m, you should get the following output in command line, and a plot will be shown by matlab.

```
>> Ch9
d1 =
         0         0         0
   54.6600         0         0
   79.0600   28.4800         0

d2 =
         0         0         0
   32.0828         0         0
   47.5451   16.2049         0

dinf =
         0         0         0
   27.9800         0         0
   40.9000   12.9200         0

dB =
         0         0         0
   26.5214         0         0
   49.9629    3.9286         0

SB + SW = ST verified.
```

In the report, you should

1. Include the generated plot in your report, observe and report the separability of the three classes.
2. Observe distance matrices $d_1, d_2, d_\infty$ and $d_B$. Compare and sort the distances between class 1, 2 and 3.
3. Report SB, SW and ST matrix in the way like $SB = [\cdots], SW = [\cdots]$ and $ST = [\cdots]$.

### 1.2 Ch10.m

After you finish Ch10.m, you should get the following output in command line,

```
>> Ch10
dB =
     0     0     0     0     0     0     0     0     0     0
   NaN     0     0     0     0     0     0     0     0     0
   NaN   NaN     0     0     0     0     0     0     0     0
   NaN   NaN   NaN     0     0     0     0     0     0     0
   NaN   NaN   NaN   NaN     0     0     0     0     0     0
   NaN   NaN   NaN   NaN   NaN     0     0     0     0     0
   NaN   NaN   NaN   NaN   NaN   NaN     0     0     0     0
```

```
NaN  NaN  NaN  NaN  NaN  NaN  NaN    0    0    0
NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN    0    0
NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN    0
```

In the report, you should

1. Include the generated energy distribution plot, observe and analyze the differences between them.
2. Include the generated plots of handwritten digit dataset in 2-D space and 3-D space. Observe and report their separability.
3. Include the generated digits plot composed of 16 subplots. The $i$'th subplot keeps the first $i * 16$ eigenvalues, and the last one is the original digit. Determine how many eigenvalues we need to keep the original digit distinguishable.

# Notes

1. These Two problems should be included in a single report with headings.
2. Source code and report should be compressed into a single .zip file named **Group_xx.zip** and handed on the canvas before **next Monday midnight, July 17 23:59**.