

EECS 545 Homework 2

Yuan Yin

September 26, 2018

```
import numpy as np

# Process the data
z = np.genfromtxt('spambase.data', dtype = float, delimiter = ',')
np.random.seed(0) # Seed the random number generator
rp = np.random.permutation(z.shape[0]) # random permutation of indices
z = z[rp,:] # shuffle the rows of z
x = z[:, :-1]
y = z[:, -1]
x_train = x[0:2000,:]
y_train = y[0:2000]
x_test = x[2000:x.shape[0],:]
y_test = y[2000:y.shape[0]]

# Quantize variables with option 1 where values equal to the median to 2
mid = np.median(x,axis=0)
row_num = x.shape[0]
col_num = x.shape[1]
x1 = x
for i in range(row_num):
    for j in range(col_num):
        if x1[i,j] >= mid[j]:
            x1[i,j] = 2
        else:
            x1[i,j] = 1

# Quantize variables with option 2 where values equal to the median to 1
mid = np.median(x,axis=0)
row_num = x.shape[0]
col_num = x.shape[1]
x2 = x
for i in range(row_num):
    for j in range(col_num):
        if x2[i,j] >= mid[j]:
            x2[i,j] = 2
        else:
            x2[i,j] = 1
```
