

Digits 3D

Practical assignment

PATTERN RECOGNITION COURSE FALL 2019

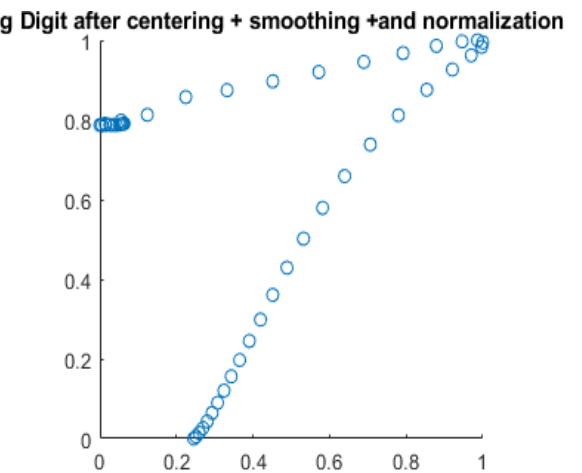
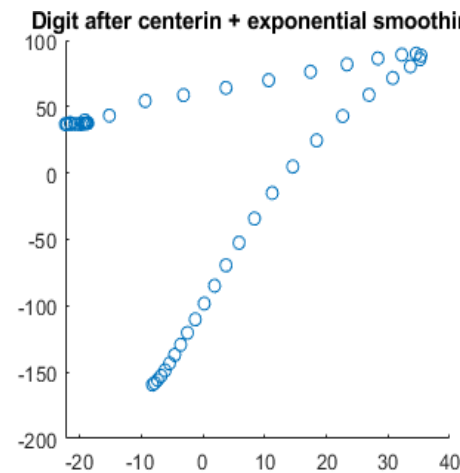
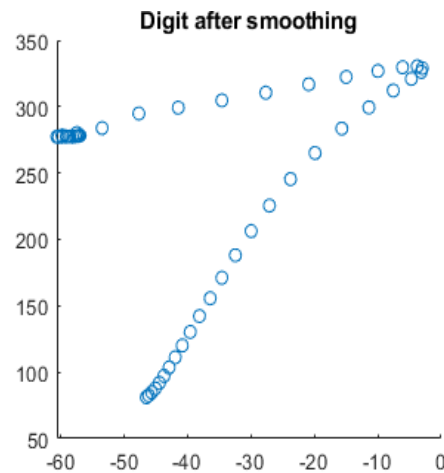
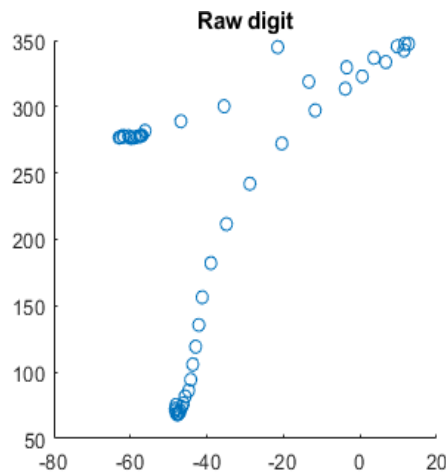
JONI KETTUNEN

I'm at work during the seminar, u
can scroll through this if u want

Data preprocessing

Pre-processing steps

1. Gather data from each file into large struct, remove Z (depth) axis
2. Exponential smoothing on each digit
3. Data centering
4. Min-max normalization



Feature extraction

- Transform time series data into 12*12 'pixel' matrix. If timeseries goes through the pixel the pixel is activated as 1. Pixel has value 0 otherwise
 - Transform each pixel to one column in final dataframe
- Calculate extra features and add into the final dataframe →

Feature	Explanation
startX	Start position X on the time series data
startY	Start Y position on the time series data
afterstartX	Position X of the timeseries data, when 33% of the datapoints are gone through
afterstartY	Position Y of the timeseries data, when 33% of the datapoints are gone through
beforeendX	Position X of the timeseries data, when 66% of the datapoints are gone through
beforeendY	Position Y of the timeseries data, when 66% of the datapoints are gone through
endX	End position X on the time series data
endY	End position Y on the time series data
startDirX	Direction of the time series on x axis at starting point
startDirY	Direction of the time series on Y axis at starting point
afterstartDirX	Direction of the time series on X axis, 33% of the timeseries values are gone through
afterstartDirY	Direction of the time series on Y axis, 33% of the timeseries values are gone through
beforeendX	Direction of the time series on X axis, 66% of the timeseries values are gone through
beforeendY	Direction of the time series on Y axis, 66% of the timeseries values are gone through
Eigenvector11	Highest Variability direction of scaled + smoothed time series <u>data</u> (no normalization)
Eigenvector12	Highest Variability direction of scaled + smoothed time series <u>data</u> (no normalization)
Eigenvector21	Second highest Variability direction of scaled + smoothed time series <u>data</u> (no normalization)
Eigenvector22	Second highest Variability direction of scaled + smoothed time series <u>data</u> (no normalization)
Eigenvalue1	Variability amount on highest variability direction
Eigenvalue2	Variability amount on second highest variability direction

Classification model

- Multilayer perceptron
 - Neural network with two layers, each layer has 10 nodes
 - 97.51% accuracy in test set (with 200 digits in test set)



Classification accuracy 97.51%
NN layers: 2
10 neurons in layer 1
10 neurons in layer 2
Epoch error stop limit: 7

Misclassified digits

- 5 digits out of 200 in test set were misclassified

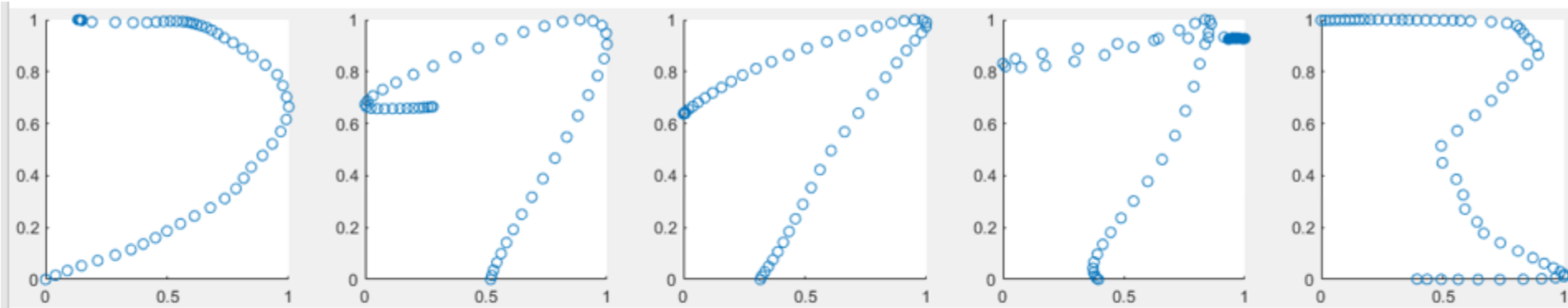


Figure 5 Misclassified digits (pre-processed)

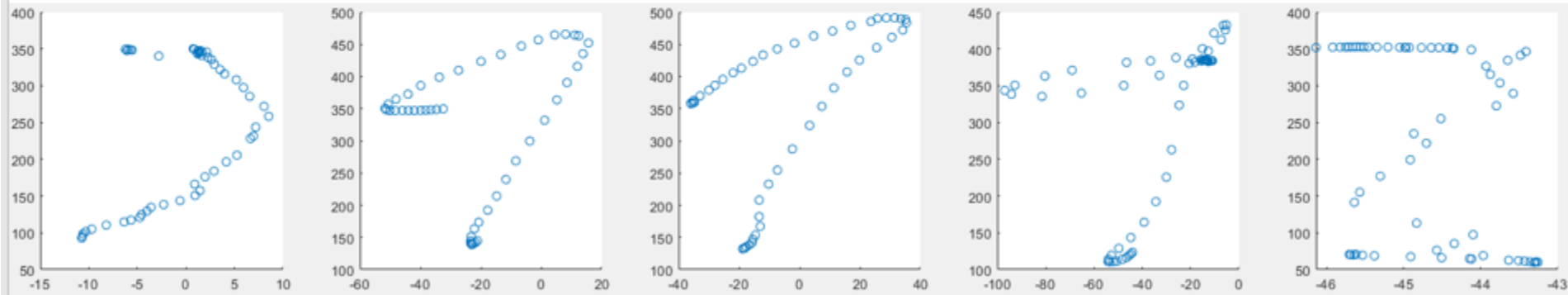


Figure 6 Misclassified digits (raw-digits)

Classification performance

- Classification performance, 500 train digits and 500 test digits.

		Actual									
		0	1	2	3	4	5	6	7	8	9
Predicted	0	50	0	0	0	0	0	1	0	0	0
	1	0	39	0	1	0	0	0	0	0	0
	2	0	2	47	0	0	0	0	0	0	0
	3	0	3	0	49	0	0	0	0	0	1
	4	0	0	0	0	47	0	0	0	0	0
	5	0	1	0	0	0	50	0	0	0	1
	6	0	0	1	0	0	0	49	0	0	0
	7	0	3	1	0	0	0	0	50	0	1
	8	0	0	1	0	0	0	0	0	50	0
	9	0	2	0	0	3	0	0	0	0	48

Figure 7 Confusion matrix done with 50% of the data in testing set