EE40098

Coursework C: Automated Feedback

Submission downloaded 12/12/2023 at 13:58.
Feedback generated on Tuesday 12th December, 2023 at 14:22.

George Madeley

University of Bath

Department of Electronic and Electrical Engineering

Formative Feedback Only

All feedback contained within this document is purely indicative of your current solutions performance. Final grades for this assignment will consider a range of additional factors when producing a mark, meaning there may be some slight change between the performance seen here and your final grades. This feedback is intended to provide you with some insight into the performance of your approach prior to submission, allowing you to further improve your solution as the semester progresses.

1   Dataset D2

1.1   General Data Properties

Table 1: Properties of the provided dataset, D2, provided for reference. These may also be found in the coursework specification.

Property	Value
SNR (Approx.)	60 dB
Num. of Classes	5
Classification Status	Unlabelled
Sampling Rate	25 kHz

Table 2: The D2.mat submission file content metadata. As outlined in the coursework specification, this should include an Index and Class vector. If you have labelled every identified pulse, you should expect the number of class entries to match the number of index entries in your submission data.

Name	Data Type	Data Shape
Class	<class 'numpy.ndarray'>	(1, 1840)
Index	<class 'numpy.ndarray'>	(1, 1840)

1.2   Submission Performance

The index identification accuracy (i.e. how well you identified where pulses were in the data) was first assessed before class identification was considered. This yielded the following metrics:

Table 3: Index identification performance for dataset D2, reported as precision and recall. In both cases, a value closer to 1.0 is better.

Metric	Value
Precision	0.965761
Recall	0.80553

The class identification was then considered, resulting in the following confusion matrix. Please note, any missed pulses will not be present in this result - so while it may look good, you will also need to consider the results found above in the index performance table.

Predicted Class

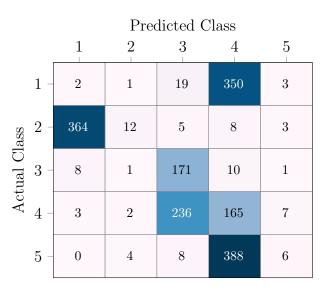|	1	2	3	4	5
1	2	1	19	350	3
2	364	12	5	8	3
3	8	1	171	10	1
4	3	2	236	165	7
5	0	4	8	388	6

Actual Class

Figure 1: Confusion matrix for the dataset D2 submission file. A good result should have most values sitting along the lead diagonal. The raw count has been used here instead of percentage to provide you with additional insight.

2   Dataset D3

2.1   General Data Properties

Table 4: Properties of the provided dataset, D3, provided for reference. These may also be found in the coursework specification.

Property	Value
SNR (Approx.)	40 dB
Num. of Classes	5
Classification Status	Unlabelled
Sampling Rate	25 kHz

Table 5: The D3.mat submission file content metadata. As outlined in the coursework specification, this should include an Index and Class vector. If you have labelled every identified pulse, you should expect the number of class entries to match the number of index entries in your submission data.

Name	Data Type	Data Shape
Class	<class 'numpy.ndarray'>	(1, 1944)
Index	<class 'numpy.ndarray'>	(1, 1944)

2.2   Submission Performance

The index identification accuracy (i.e. how well you identified where pulses were in the data) was first assessed before class identification was considered. This yielded the following metrics:

Table 6: Index identification performance for dataset D3, reported as precision and recall. In both cases, a value closer to 1.0 is better.

Metric	Value
Precision	0.950103
Recall	0.904949

The class identification was then considered, resulting in the following confusion matrix. Please note, any missed pulses will not be present in this result - so while it may look good, you will also need to consider the results found above in the index performance table.

Predicted Class

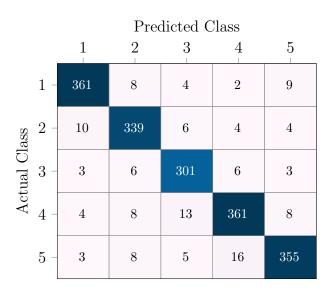|	1	2	3	4	5
1	361	8	4	2	9
2	10	339	6	4	4
3	3	6	301	6	3
4	4	8	13	361	8
5	3	8	5	16	355

Actual Class

Figure 2: Confusion matrix for the dataset D3 submission file. A good result should have most values sitting along the lead diagonal. The raw count has been used here instead of percentage to provide you with additional insight.

3   Dataset D4

3.1   General Data Properties

Table 7: Properties of the provided dataset, D4, provided for reference. These may also be found in the coursework specification.

Property	Value
SNR (Approx.)	20 dB
Num. of Classes	5
Classification Status	Unlabelled
Sampling Rate	25 kHz

Table 8: The D4.mat submission file content metadata. As outlined in the coursework specification, this should include an Index and Class vector. If you have labelled every identified pulse, you should expect the number of class entries to match the number of index entries in your submission data.

Name	Data Type	Data Shape
Class	<class 'numpy.ndarray'>	(1, 2539)
Index	<class 'numpy.ndarray'>	(1, 2539)

3.2   Submission Performance

The index identification accuracy (i.e. how well you identified where pulses were in the data) was first assessed before class identification was considered. This yielded the following metrics:

Table 9: Index identification performance for dataset D4, reported as precision and recall. In both cases, a value closer to 1.0 is better.

Metric	Value
Precision	0.859393
Recall	0.849688

The class identification was then considered, resulting in the following confusion matrix. Please note, any missed pulses will not be present in this result - so while it may look good, you will also need to consider the results found above in the index performance table.
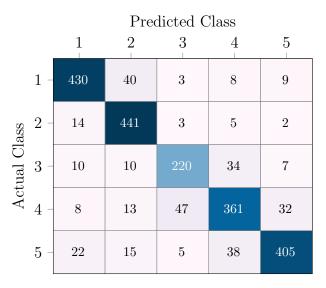
Figure 3: Confusion matrix for the dataset D4 submission file. A good result should have most values sitting along the lead diagonal. The raw count has been used here instead of percentage to provide you with additional insight.

4 Dataset D5

4.1 General Data Properties

Table 10: Properties of the provided dataset, D5, provided for reference. These may also be found in the coursework specification.

Property	Value
SNR (Approx.)	0 dB
Num. of Classes	5
Classification Status	Unlabelled
Sampling Rate	25 kHz

Table 11: The D5.mat submission file content metadata. As outlined in the coursework specification, this should include an Index and Class vector. If you have labelled every identified pulse, you should expect the number of class entries to match the number of index entries in your submission data.

Name	Data Type	Data Shape
Class	<class 'numpy.ndarray'>	(1, 2389)
Index	<class 'numpy.ndarray'>	(1, 2389)

4.2 Submission Performance

The index identification accuracy (i.e. how well you identified where pulses were in the data) was first assessed before class identification was considered. This yielded the following metrics:

Table 12: Index identification performance for dataset D5, reported as precision and recall. In both cases, a value closer to 1.0 is better.

Metric	Value
Precision	0.678108
Recall	0.795287

The class identification was then considered, resulting in the following confusion matrix. Please note, any missed pulses will not be present in this result - so while it may look good, you will also need to consider the results found above in the index performance table.
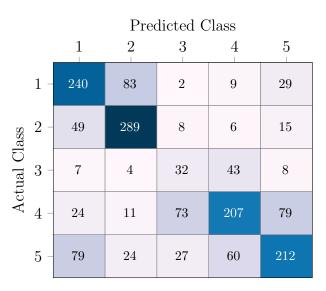
Figure 4: Confusion matrix for the dataset D5 submission file. A good result should have most values sitting along the lead diagonal. The raw count has been used here instead of percentage to provide you with additional insight.

5   Dataset D6

5.1   General Data Properties

Table 13: Properties of the provided dataset, D6, provided for reference. These may also be found in the coursework specification.

Property	Value
SNR (Approx.)	<0 dB
Num. of Classes	5
Classification Status	Unlabelled
Sampling Rate	25 kHz

Table 14: The D6.mat submission file content metadata. As outlined in the coursework specification, this should include an Index and Class vector. If you have labelled every identified pulse, you should expect the number of class entries to match the number of index entries in your submission data.

Name	Data Type	Data Shape
Class	<class 'numpy.ndarray'>	(1, 3206)
Index	<class 'numpy.ndarray'>	(1, 3206)

5.2   Submission Performance

The index identification accuracy (i.e. how well you identified where pulses were in the data) was first assessed before class identification was considered. This yielded the following metrics:

Table 15: Index identification performance for dataset D6, reported as precision and recall. In both cases, a value closer to 1.0 is better.

Metric	Value
Precision	0.670306
Recall	0.711825

The class identification was then considered, resulting in the following confusion matrix. Please note, any missed pulses will not be present in this result - so while it may look good, you will also need to consider the results found above in the index performance table.

Predicted Class

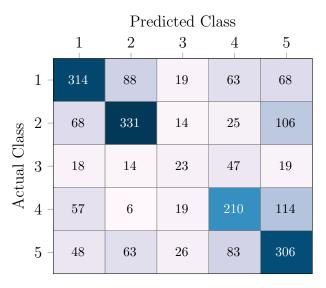|	1	2	3	4	5
1	314	88	19	63	68
2	68	331	14	25	106
3	18	14	23	47	19
4	57	6	19	210	114
5	48	63	26	83	306

Figure 5: Confusion matrix for the dataset D6 submission file. A good result should have most values sitting along the lead diagonal. The raw count has been used here instead of percentage to provide you with additional insight.