K-Nearest Neighbor Based Spoken Letter Recognition

Annika Shankwitz

Department of Linguistics

Spoken Letter Recognition

- Hearing a letter pronounced → Identifying the written letter

Spoken Letter Recognition

- Hearing a letter pronounced → Identifying the written letter

“B”

B

ISOLET – the Isolated Letter Speech Recognition Database

- Contains isolated spoken letters of the English alphabet
  - 150 speakers, 2 productions of the alphabet per speaker
  - 7,800 instances
  - 617 acoustical features

[1, 2]

ISOLET – the Isolated Letter Speech Recognition Database

- Contains isolated spoken letters of the English alphabet
  - 150 speakers, 2 productions of the alphabet per speaker
  - 7,800 instances
  - 617 acoustical features

- Has 5 parts: ISOLET1-5
  - Each part contains 30 speakers, 15 male and 15 female
  - Training: ISOLET1-4 (80%)
  - Testing: ISOLET5 (20%)

[1, 2]

ISOLET – Previous Work

- Neural network – 95.5% accuracy
- Hidden Markov Model – 96.6% accuracy

[3, 6]

ISOLET – Previous Work

- Neural network – 95.5% accuracy
- Hidden Markov Model – 96.6% accuracy

- International Olympic Committee algorithm – KNN algorithm designed for high-dimensional, multi-class classification tasks
  - ↑ Only KNN approach applied to ISOLET

[3, 6, 4]

Research Questions

Q1: Does a KNN model perform well on ISOLET?

↑ with techniques to reduce data dimensionality and increase model efficiency

Q2: Which spoken letters are easiest for the KNN model to classify?

Q3: Which spoken letters are hardest for the KNN model to classify?

Methodology – Data Preprocessing

- Min-Max scaling

Methodology – Data Preprocessing

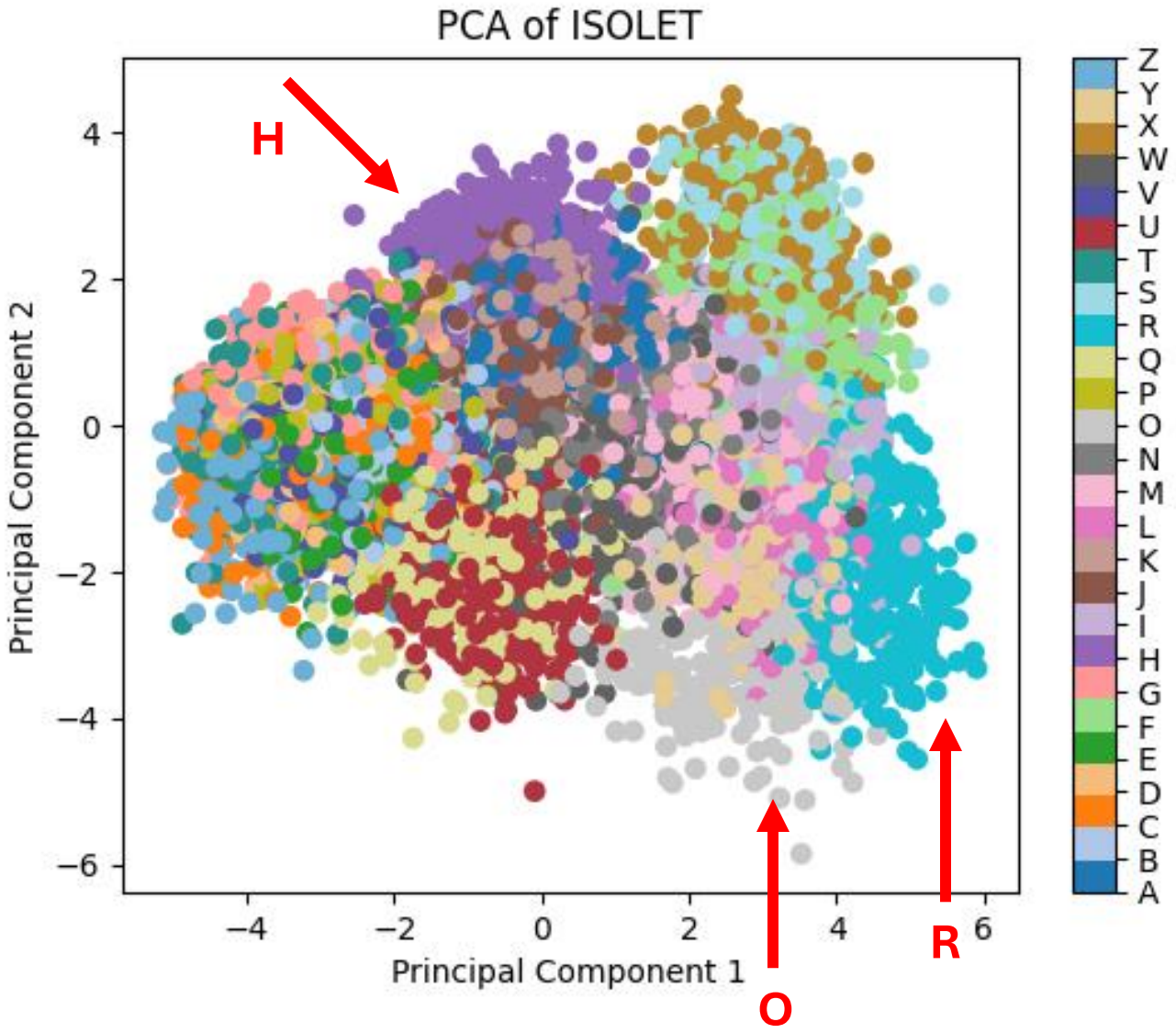
- Min-Max scaling

- PCA
  - 617 dimensions → 179 dimensions
    (while capturing 95% of variance)

Methodology – Data Preprocessing

- Min-Max scaling

- PCA
  - 617 dimensions → 179 dimensions (while capturing 95% of variance)
  - Some classes form distinct clusters
    - H, R, O
  - Most classes don't form distinct clusters



Methodology – the KNN model

- Scikit-learn’s KNeighborsClassifier()
  - gridSearchCV()
    - k = [1,3,5,7,9,11,13,15,17,19,21]
    - weights= uniform vs distance
    - algorithm = ball tree vs KD tree
  - 5-fold cross validation

Methodology – the KNN model

- Scikit-learn’s KNeighborsClassifier()
  - gridSearchCV()
    - k = [1,3,5,7,9,11,13,15,17,19,21]
    - weights= uniform vs distance
    - algorithm = ball tree vs KD tree
  - 5-fold cross validation

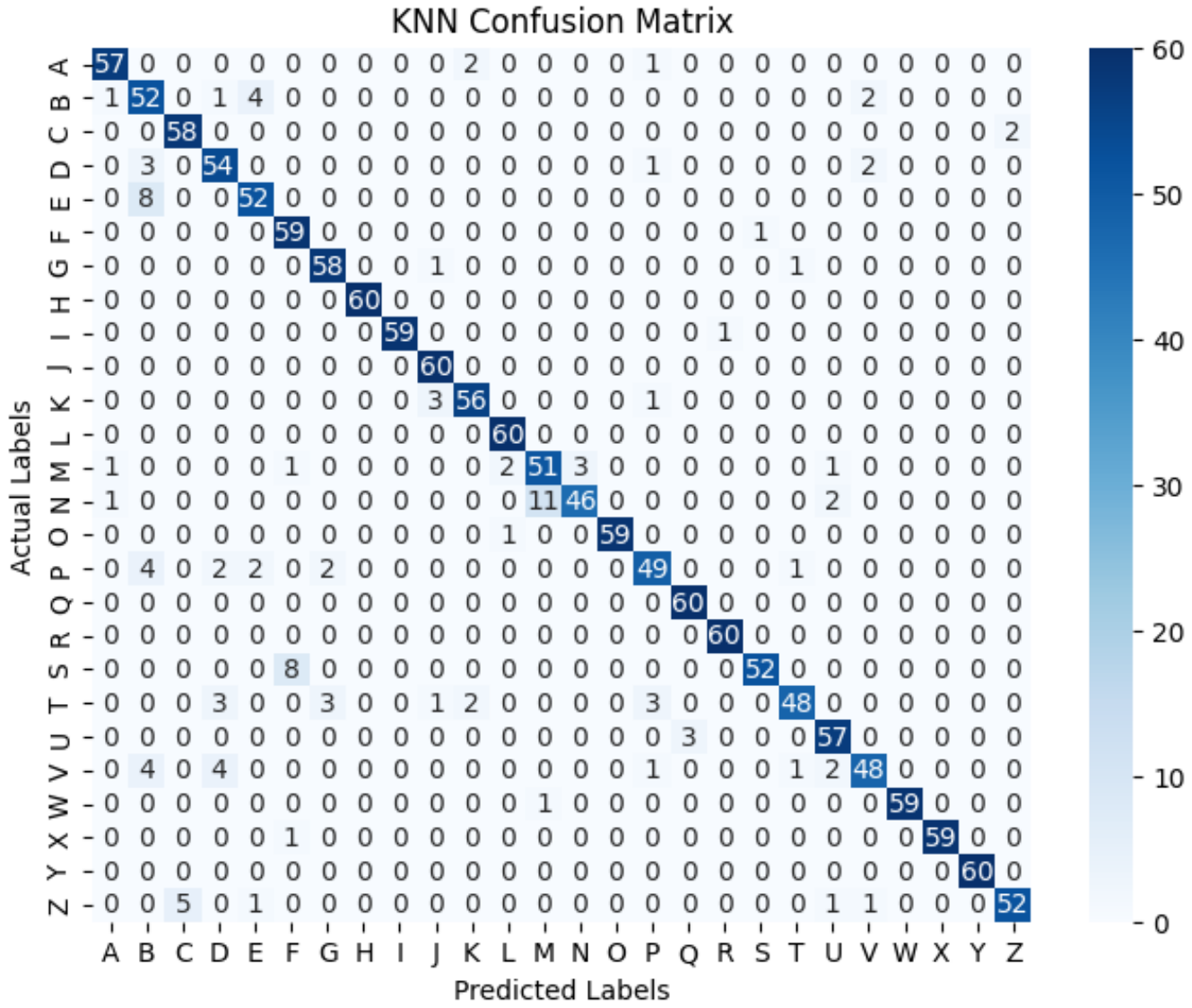
- Best Parameters
  - k = 9
  - weights = distance
  - algorithm = ball tree

[7]

Q1: Does a KNN model perform well on ISOLET?

Yes!

- Macro-Precision: 0.93
- Macro-Recall: 0.93
- Macro-F1 score: 0.93
- Accuracy: 0.93



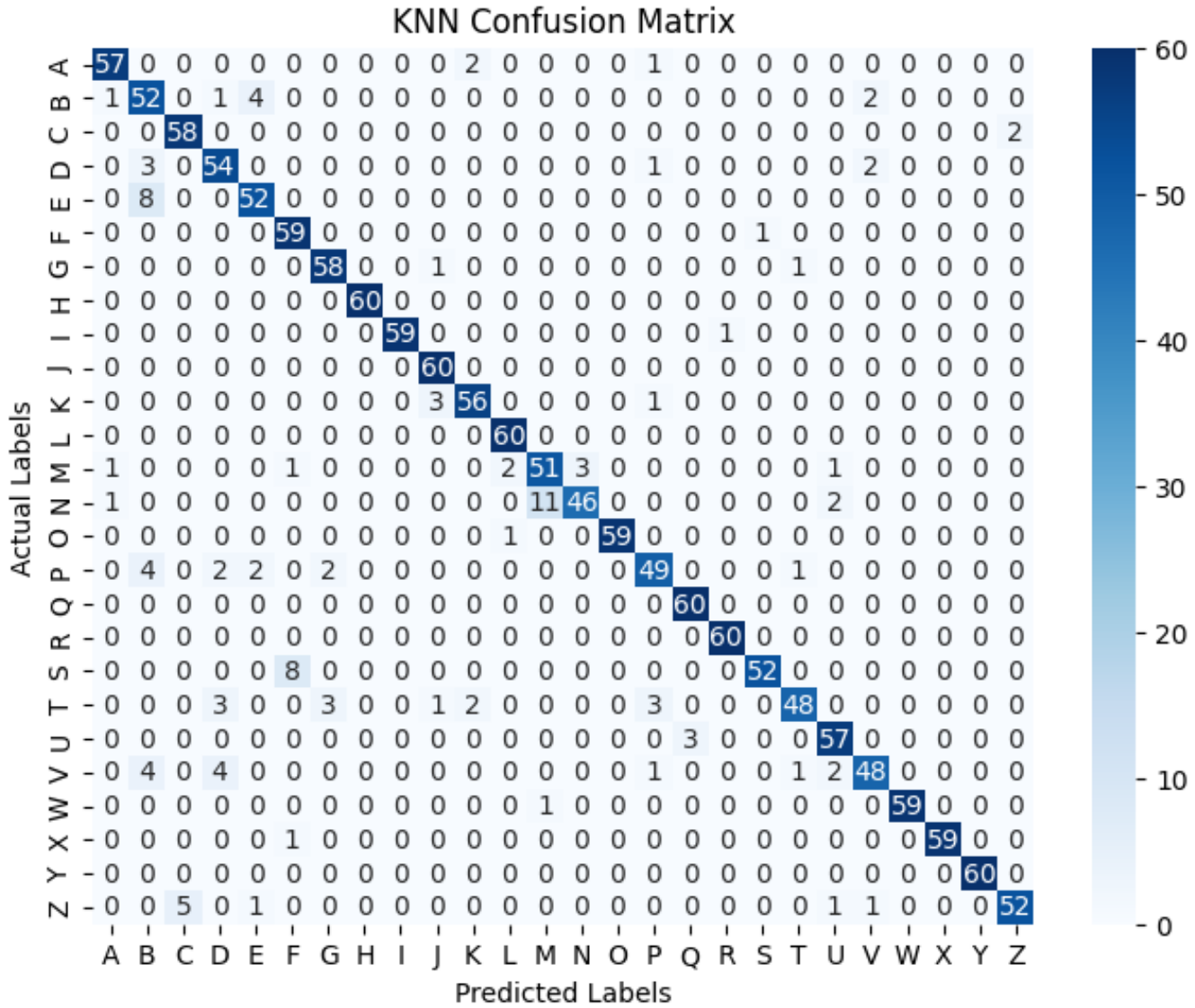
Q1: Does a KNN model perform well on ISOLET?

Yes!

- Macro-Precision: 0.93
- Macro-Recall: 0.93
- Macro-F1 score: 0.93
- Accuracy: 0.93

Compared to Previous Work:

- 2.5% less accurate than a neural network
- 3.6% less accurate than a Hidden Markov Model
  - ↑ both more complicated models



[3, 6]

Q2: Which spoken letters are easiest for the KNN model to classify?

- H (F1 score 1.0)
- I (F1 score 0.99)
- O (F1 score 0.99)
- R (F1 score 0.99)
- W (F1 score 0.99)
- X (F1 score 0.99)
- Y (F1 score 1.0)

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q2: Which spoken letters are easiest for the KNN model to classify?

- H (F1 score 1.0)
- I (F1 score 0.99)
- O (F1 score 0.99)
- R (F1 score 0.99)
- W (F1 score 0.99)
- X (F1 score 0.99)
- Y (F1 score 1.0)

- All have distinct pronunciations

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q3: Which spoken letters are hardest for the KNN model to classify?

- B (F1 score 0.79)
- M (F1 score 0.84)
- N (F1 score 0.84)
- P (F1 score 0.84)
- V (F1 score 0.85)

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q3: Which spoken letters are hardest for the KNN model to classify?

- B (F1 score 0.79)
- M (F1 score 0.84)
- N (F1 score 0.84)
- P (F1 score 0.84)
- V (F1 score 0.85)

- All do not have distinct pronunciations

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q3: Which spoken letters are hardest for the KNN model to classify?

- B (F1 score 0.79)
- M (F1 score 0.84)
- N (F1 score 0.84)
- P (F1 score 0.84)
- V (F1 score 0.85)

- All do not have distinct pronunciations
  - M is very similar to N

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q3: Which spoken letters are hardest for the KNN model to classify?

- B (F1 score 0.79)
- M (F1 score 0.84)
- N (F1 score 0.84)
- P (F1 score 0.84)
- V (F1 score 0.85)

- All do not have distinct pronunciations
  - M is very similar to N
  - B, P, V are all similar to each other, and to D, E, T

IPA TRANSCRIPTIONS OF THE ENGLISH ALPHABET

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Q3: Which spoken letters are hardest for the KNN model to classify?

- B (F1 score 0.79)
- M (F1 score 0.84)
- N (F1 score 0.84)
- P (F1 score 0.84)
- V (F1 score 0.85)

- All do not have distinct pronunciations
  - M is very similar to N
  - B, P, V are all similar to each other, and to D, E, T
  - Next lowest letters: T (F1 score=0.86), D (F1 score=0.87), and E (F1 score=0.87)

IPA Transcriptions of the English Alphabet

Letter	IPA Transcription	Letter	IPA Transcription
A	[eɪ]	N	[ɛn]
B	[bi]	O	[ow]
C	[si]	P	[pi]
D	[di]	Q	[kju]
E	[i]	R	[aɪ]
F	[ɛf]	S	[ɛs]
G	[dʒi]	T	[ti]
H	[eɪtʃ]	U	[ju]
I	[aɪ]	V	[vi]
J	[dʒeɪ]	W	[də.bl̩.ju]
K	[keɪ]	X	[ɛks]
L	[ɛl]	Y	[waɪ]
M	[ɛm]	Z	[zi]

Conclusion

Q1: Does a KNN model perform well on ISOLET?

↑ yes!

Q2: Which spoken letters are easiest for the KNN model to classify?

↑ letters with distinct pronunciations: H, I, O, R, W, X, Y

Q3: Which spoken letters are hardest for the KNN model to classify?

↑ letters without distinct pronunciations: B, M, N, P, V

References

[1] R. Cole, Y. Muthusamy, and M. Fanty. “The ISOLET spoken letter database.” 1996.

[2] M. Fanty and R. Cole. “ISOLET”. UCI Machine Learning Repository. Available: https://archive.ics.uci.edu/dataset/54/isolet.

[3] M. Fanty and R. Cole. “Spoken letter recognition”. Advances in Neural Information Processing Systems (NIPS Conference). 1990. pp. 220-226.

[4] T. Liu, K. Yang, and A. Moore. “The IOC algorithm: efficient many-class non-parametric classification for high-dimensional data”. Proceedings of the Tenth ACM SIGKDD International Conference on KnowledgeDiscovery and Data Mining. 2004.

[5] F. Markus. “Why is speech recognition difficult?”. Semantic Scholar. 2003.

[6] H. Nock and S. Young. “Loosely coupled HMMs for ASR”. Sixth International Conference on Spoken Language Processing (INTERSPEECH).2000.

[7] Pedregosa et al., “Scikit-Learn: machine learning in Python”, Journal of Machine Learning Research, vol. 12. 2011. pp. 2825-2830.