Time Series Classification

Contents

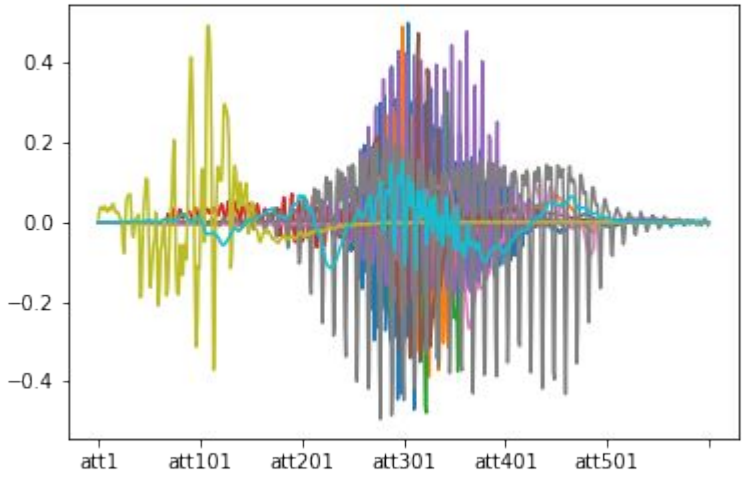
Dataset

- Sound of insect wingbeats
- Time series length: 600
- 50,000 samples
- 10 classes, for each 5,000 samples:
  - Aedes_female
  - Aedes_male
  - Fruit_flies
  - House_flies
  - Quinx_female
  - Quinx_male
  - Stigma_female
  - Stigma_male
  - Tarsalis_female
  - Tarsalis_male
- Task: classification

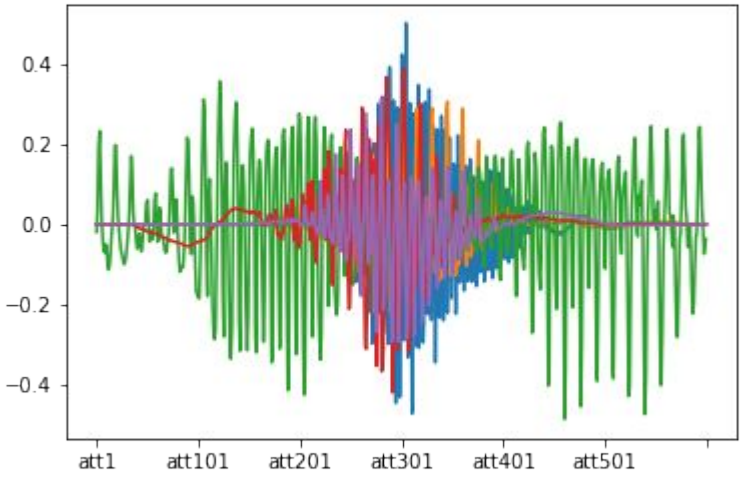
From: http://www.timeseriesclassification.com/description.php?Dataset=InsectSound

	att1	att2	att3	...	att600	target
0	0.000153	-6.578439e-05	0.000038		-0.000024	b'Aedes_female'
1	-0.000029	6.673935e-07	0.000035		0.001597	b'Aedes_female'
2	-0.019499	4.194278e-02	0.192660		-0.037126	b'Aedes_female'
3	0.000000	0.000000e+00	0.000000		0.000000	b'Aedes_female'
4	0.000023	-7.772223e-05	0.000111		0.000361	b'Aedes_female'
...	...	...	...	...	...	...
24995	-0.038172	-6.510722e-02	-0.050728		-0.043915	b'Tarsalis_male'
24996	-0.004521	-5.507769e-03	-0.003324		-0.002782	b'Tarsalis_male'
24997	0.000001	1.723497e-05	0.000005		-0.006218	b'Tarsalis_male'
24998	-0.000047	-5.119860e-05	-0.000045		-0.114638	b'Tarsalis_male'
24999	-0.011246	-1.924826e-02	-0.020571		-0.000106	b'Tarsalis_male'

Insect sound dataset.



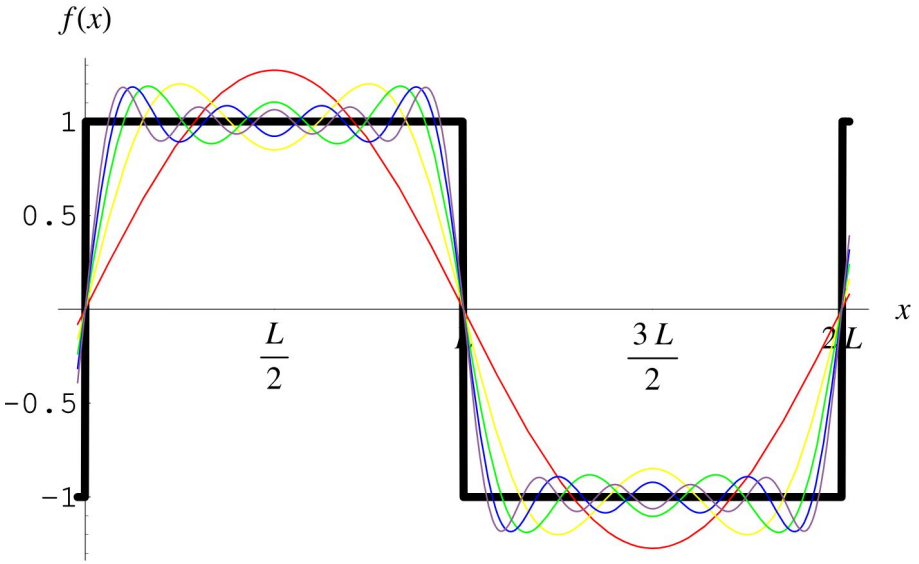
Samples of 10 individuals of different classes.



Samples of 5 individuals of class 0 (Aedes female).

Fourier Transformation

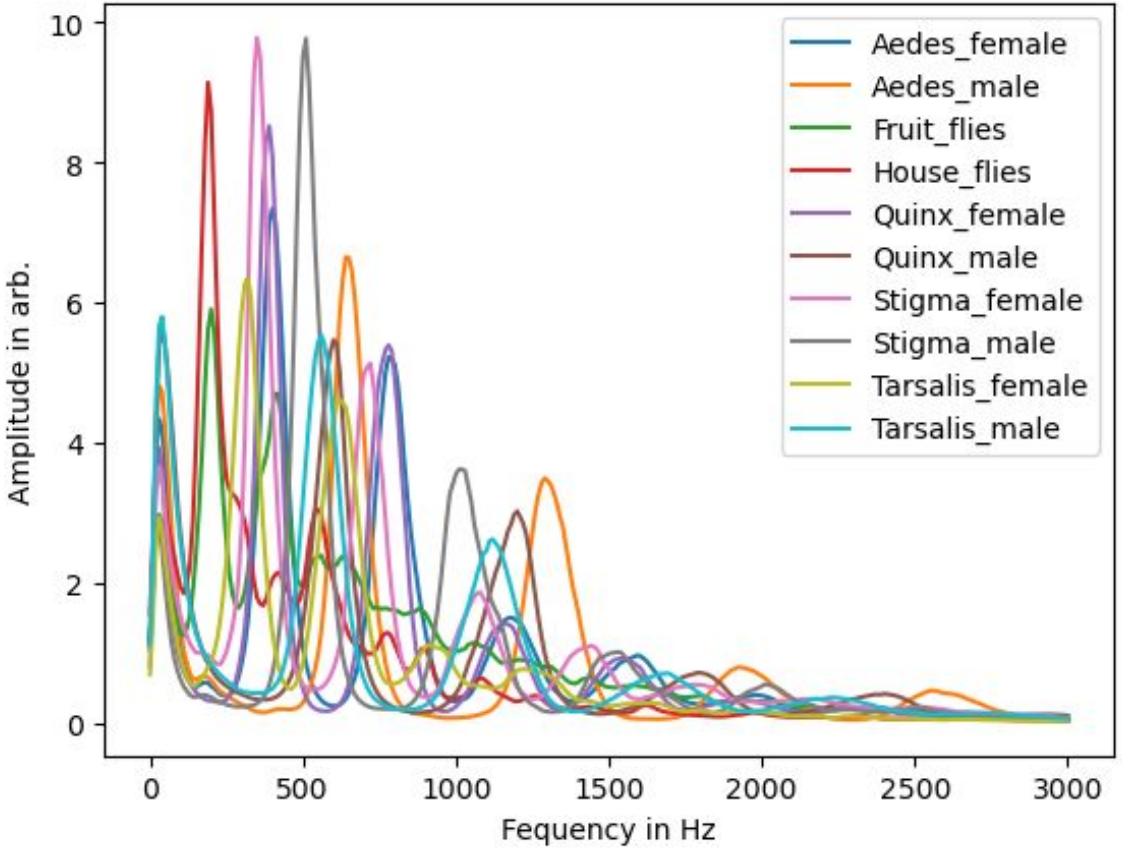
$$f(t) = \sum_{n=-\infty}^{\infty} F(n)e^{jn2\pi f_1 t}$$

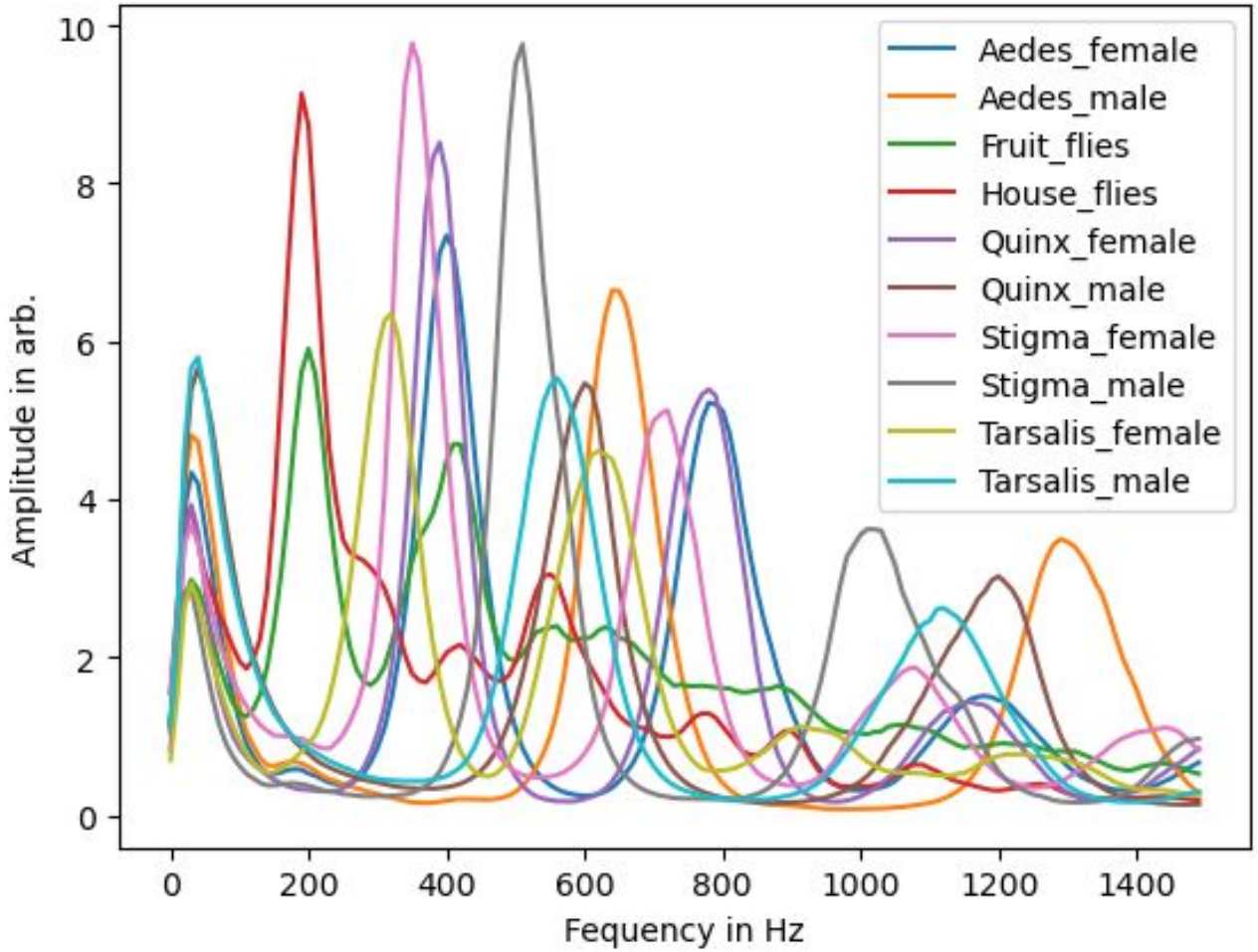


The fast Fourier transform, Brigham et al., 1967

https://mathworld.wolfram.com/FourierSeriesSquareWave.html

Frequency Spectra for all Species

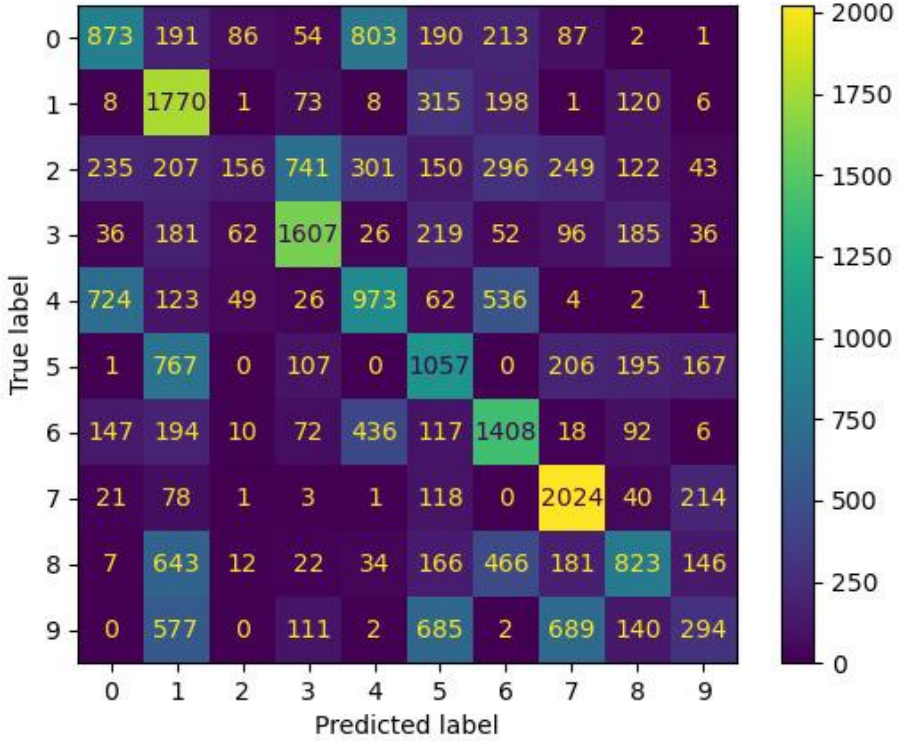



Aedes_female
Aedes_male
Fruit_flies
House_flies
Quinx_female
Quinx_male
Stigma_female
Stigma_male
Tarsalis_female
Tarsalis_male
Amplitude in arb.
Fequency in Hz
0
2
4
6
8
10
0
200
400
600
800
1000
1200
1400

Classification Algorithms

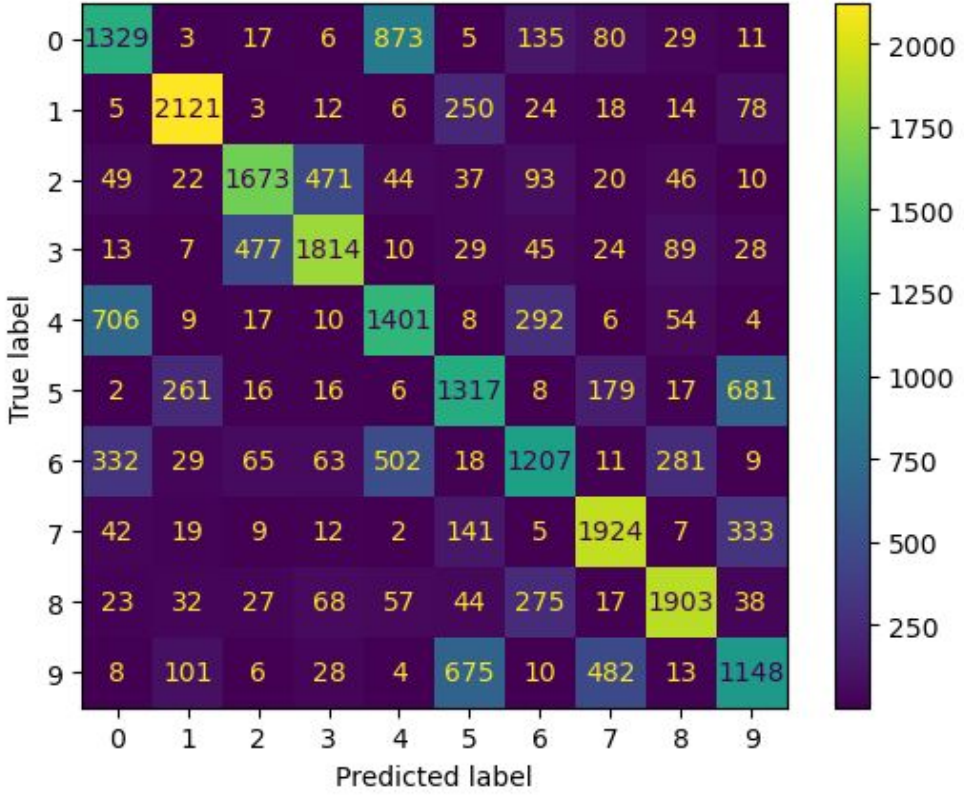
- Random guessing
- k-nearest Neighbors
  - most prominent frequency
  - entire spectrum
- Gaussian Naive Bayes
- Decision Tree
- Support Vector Classifier
- Random Forest
- Dynamic Time Warping

Most prominent Frequency and k-nearest neighbors



Accuracy score:
0.4394

k-nearest Neighbors



Accuracy score:
0.63788

Gaussian Naive Bayes

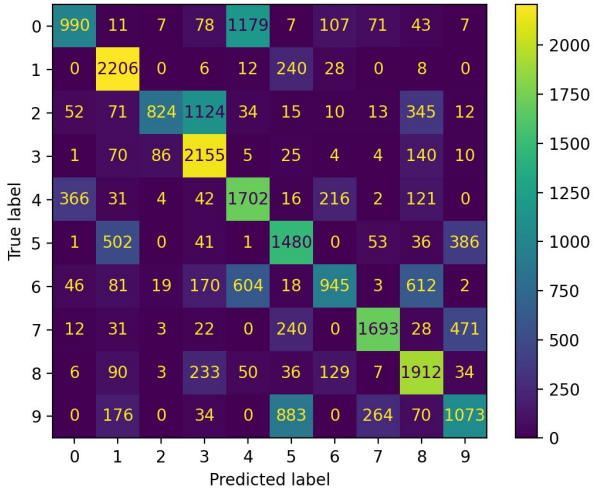
Calculates likelihood for each feature for each class

Assumption: likelihood of the features is Gaussian

Assumption: features are independent from each other

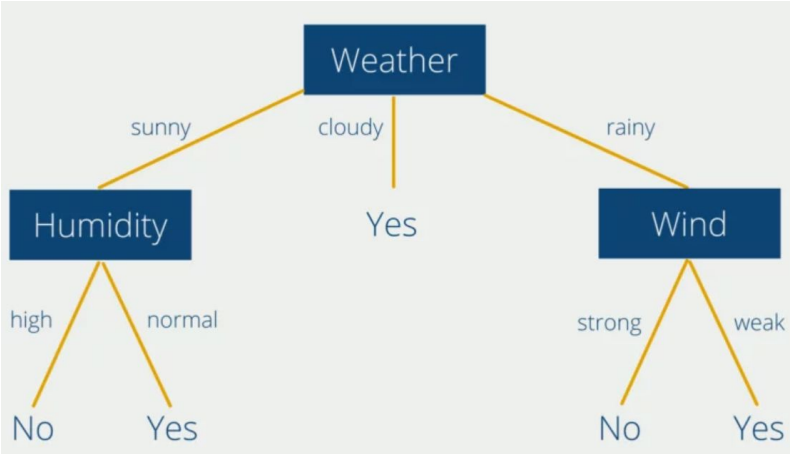
$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

From: https://scikit-learn.org/stable/modules/naive_bayes.html

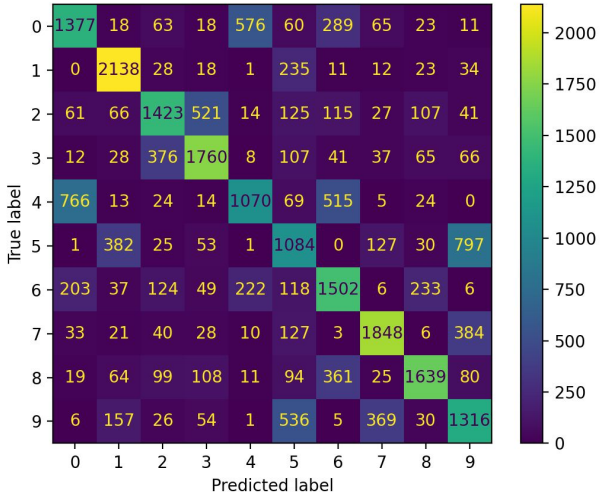


Accuracy score: 0.601413

Decision Tree

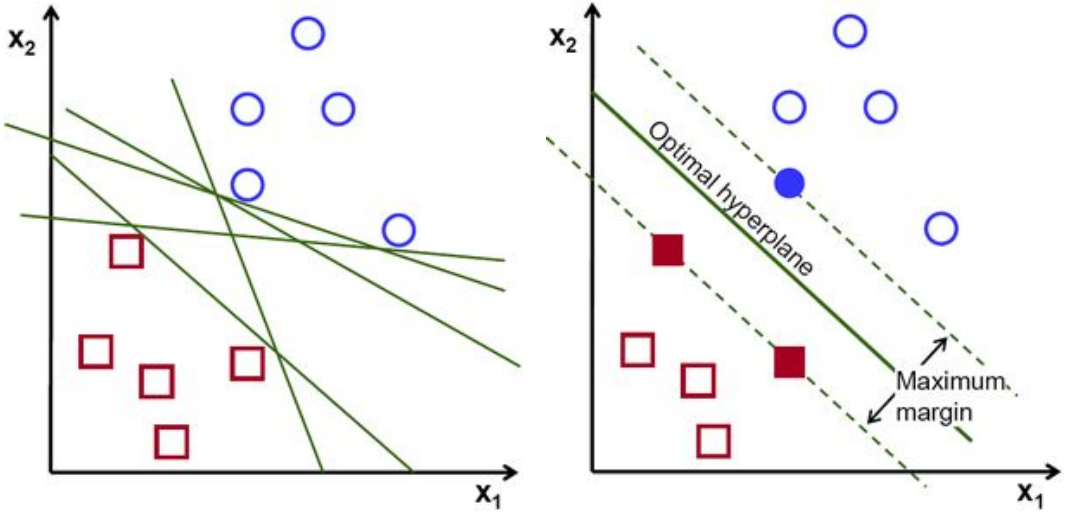


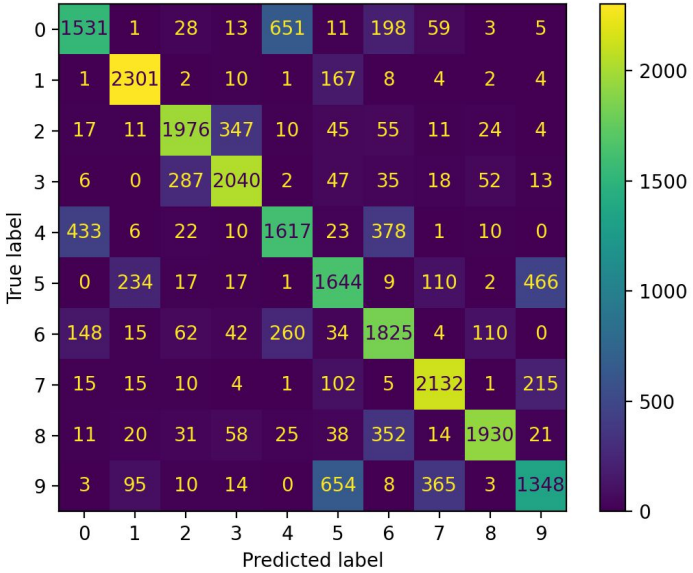
From: https://databasecamp.de/ki/decision-tree



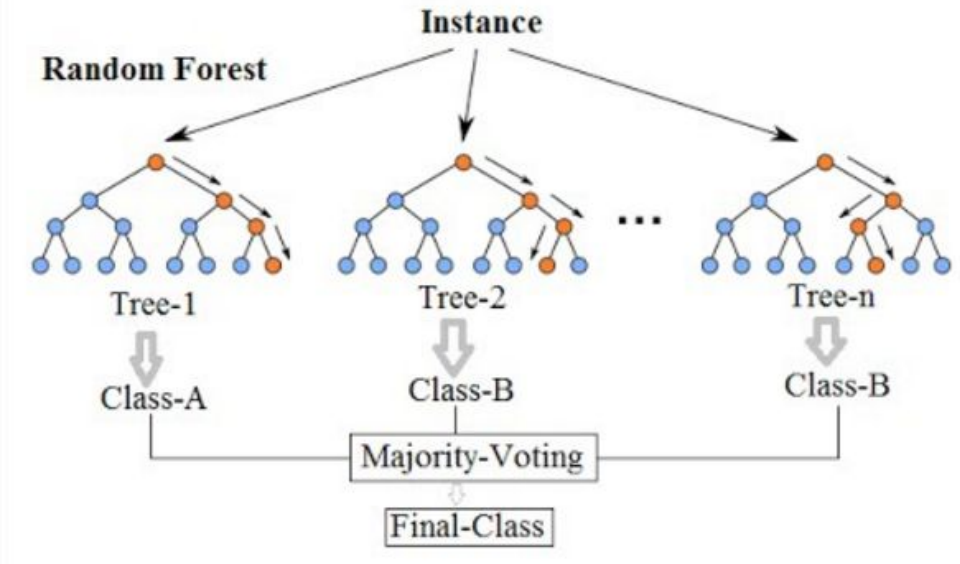
Accuracy score: 0.60864

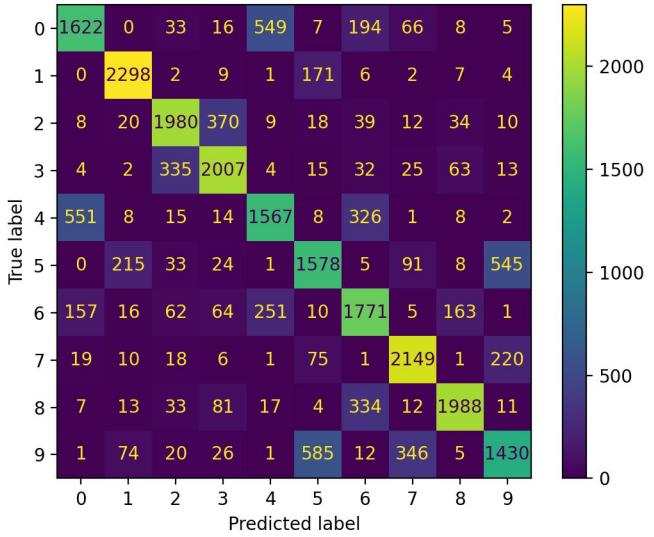
Support Vector Classifier



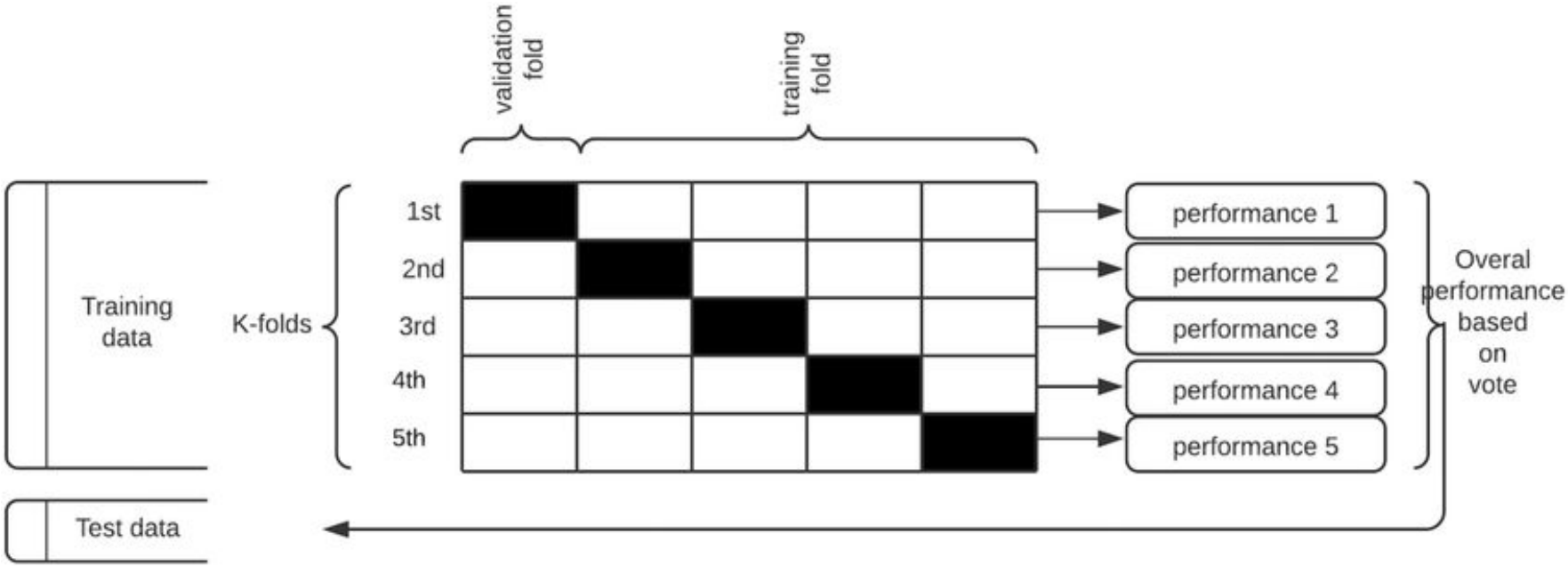


Random Forest Classifier



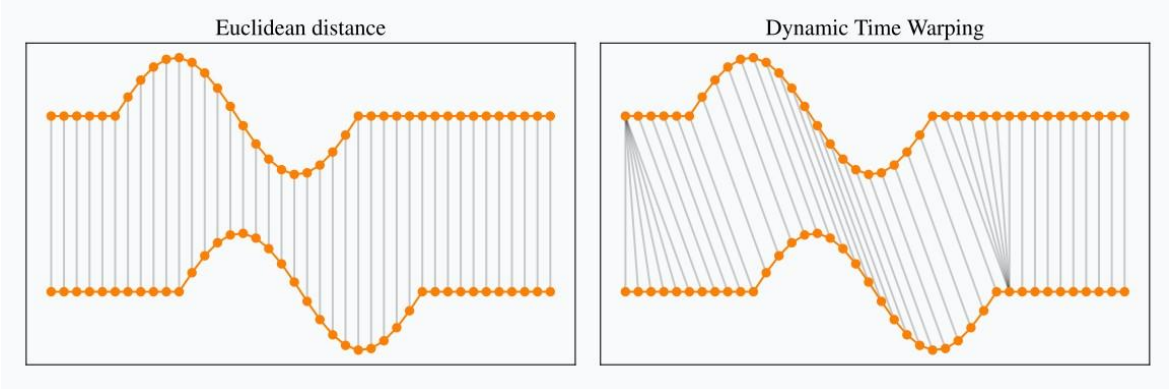


Cross Validation and Grid Search



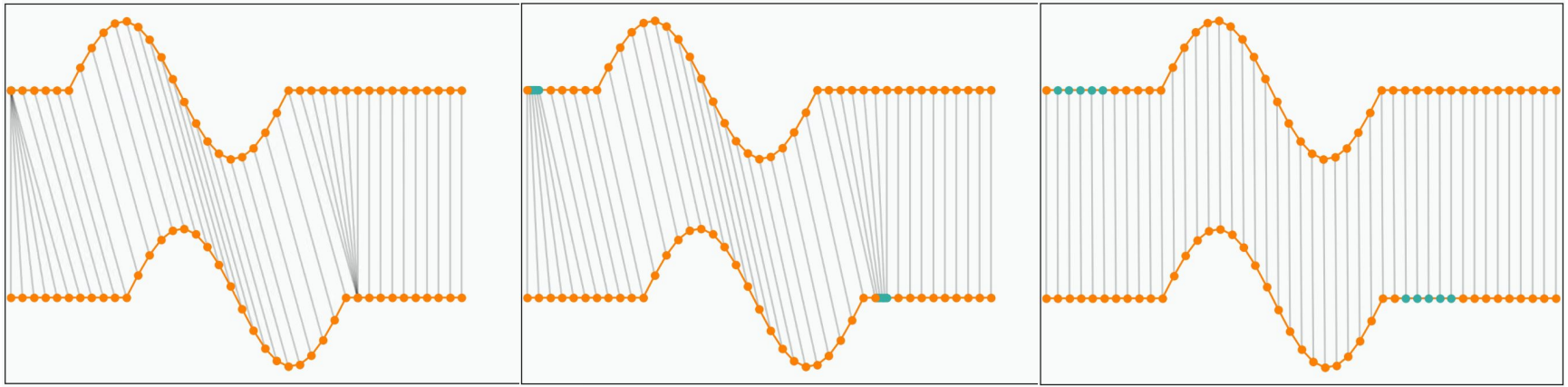
Dynamic Time Warping (DTW)

- Alignment-based metric
- Similarity: Sum of Euclidean distances between matched features
- Matched features can be at different time stamps
- Time stamps can be used for >1 alignment
- DTW used as metric in KMeans clustering



From: https://rtavenar.github.io/blog/dtw.html

$$DTW_q(x, x') = \min_{\pi \in \mathcal{A}(x,x')} \left(\sum_{(i,j)\in\pi} d(x_i, x'_j)^q \right)^{\frac{1}{q}}$$



From: https://rtavenar.github.io/blog/dtw.html

DTW - Results

- very slow
  - results were only achieved with restrictions
    - reduced dataset (5 and 50 samples for each class)
    - small number of iterations of KMeans
- bad clustering results:
  - Adjusted Random Index: 0.0015

name	accs_mean	accs_std	accs_min	accs_max	fit_time	test_time
k-Nearest-Neighbors (max frequency)	0.435120	0.003670	0.430577	0.443178	57.485401	0.876546
Decision Tree	0.596813	0.008794	0.586153	0.610944	463.918601	0.010120
Decision Tree small	0.599493	0.006254	0.586753	0.607871	229.329387	0.008096
Gaussian Naive Bayes	0.601413	0.001692	0.598872	0.603864	12.691428	0.492448
Gaussian Naive Bayes small	0.601453	0.003275	0.595392	0.607824	7.204690	0.171305
k-Nearest-Neighbors small	0.623613	0.004456	0.614545	0.629905	20.451361	1.933932
k-Nearest-Neighbors	0.628587	0.004327	0.620905	0.634825	32.327905	3.219947
Support Vector Classifier small	0.720613	0.002929	0.715829	0.724829	844.496470	31.539375
Support Vector Classifier	0.726213	0.002954	0.721502	0.730589	1399.935758	38.678565
Random Forest small	0.729293	0.004982	0.723149	0.738390	719.034351	0.530636
Random Forest	0.731787	0.004070	0.725342	0.739710	987.509642	0.546636

References

- Flying Insect Classification with Inexpensive Sensors; Yanping Chen, Adena Why, Gustavo Batista, Agenor Mafra-Neto, Eamonn Keogh, 2014
- http://www.timeseriesclassification.com/description.php?Dataset=InsectSound
- The fast Fourier transform, Brigham et al., 1967
- https://mathworld.wolfram.com/FourierSeriesSquareWave.html
- https://www.sktime.net/en/latest/get_started.html
- https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47
- https://www.researchgate.net/figure/Random-Forest-Algorithm-mechanism-Random-Forest-Algorithm-runs-into-two-stages-The_fig1_338024116
- https://scikit-learn.org/stable/modules/naive_bayes.html
- https://rtavenar.github.io/blog/dtw.html