Functional Neural Networks

Shift invariant models for functional data with applications to EEG classification

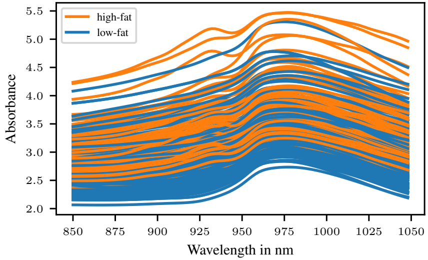
Florian Heinrichs, Mavin Heim, Corinna Weber

SNAP GmbH

July 2023

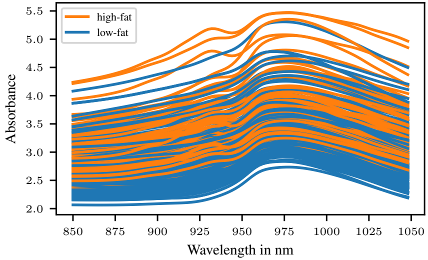
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)



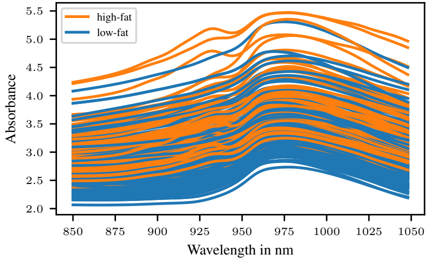
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$



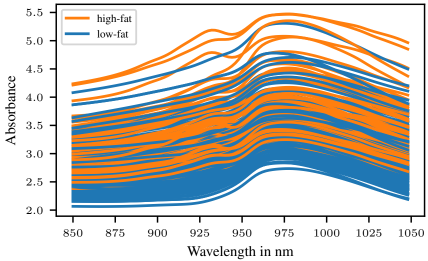
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \dots, x_{100})$
- **Does this vector representation make sense?**



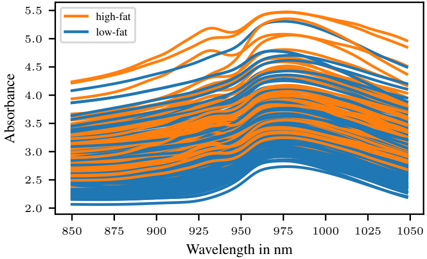
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**
- In general there is no relation between coordinates of a vector



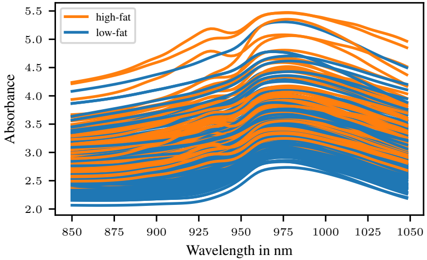
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**
- In general there is no relation between coordinates of a vector
- Information about smoothness in a vector representation is lost



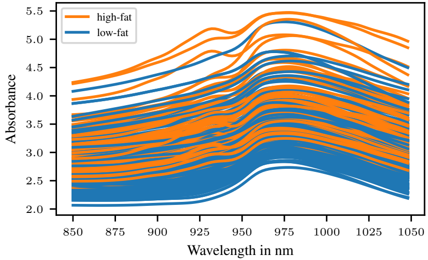
Functional Data Analysis (FDA)

- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**
- In general there is no relation between coordinates of a vector
- Information about smoothness in a vector representation is lost
- **Instead:** Consider observations as functions $x(t)$



Functional Data Analysis (FDA)

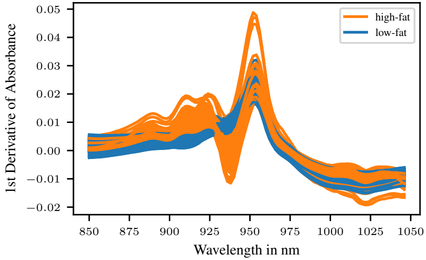
- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**
- In general there is no relation between coordinates of a vector
- Information about smoothness in a vector representation is lost
- **Instead:** Consider observations as functions $x(t)$
- Hidden information can be made visible
  - e.g. $x'(t), x''(t)$



Functional Data Analysis (FDA)

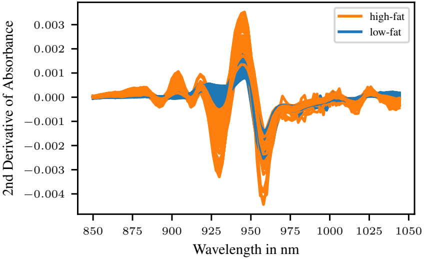
- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**

- In general there is no relation between coordinates of a vector
- Information about smoothness in a vector representation is lost
- **Instead:** Consider observations as functions $x(t)$
- Hidden information can be made visible
  - e.g. $x'(t), x''(t)$



Functional Data Analysis (FDA)

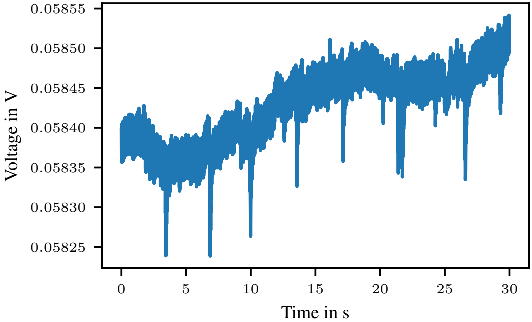
- Example: Classify meat (low-fat vs. high-fat) based on *near infrared absorbance spectrum* (measured at 100 wavelengths between 850 to 1050 nm)
- For each meat sample, we have an observation $x = (x_1, \ldots, x_{100})$
- **Does this vector representation make sense?**
- In general there is no relation between coordinates of a vector
- Information about smoothness in a vector representation is lost
- **Instead:** Consider observations as functions $x(t)$
- Hidden information can be made visible
  - e.g. $x'(t), x''(t)$



EEG Data

1 Functional Data Analysis

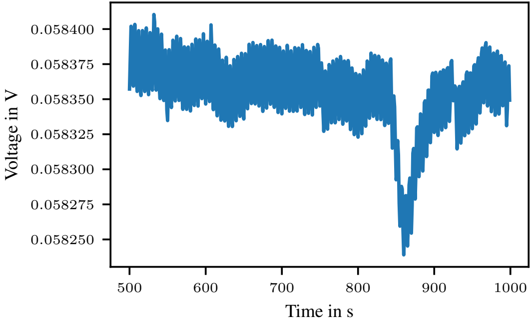
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window



EEG Data

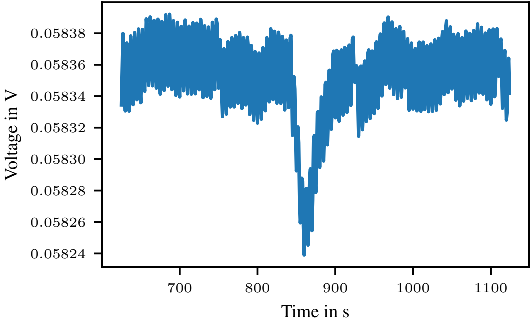
1 Functional Data Analysis

- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window



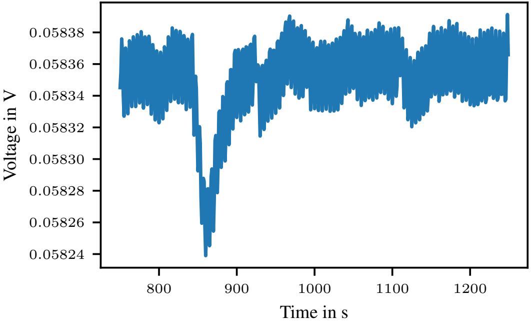
EEG Data

- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window



EEG Data

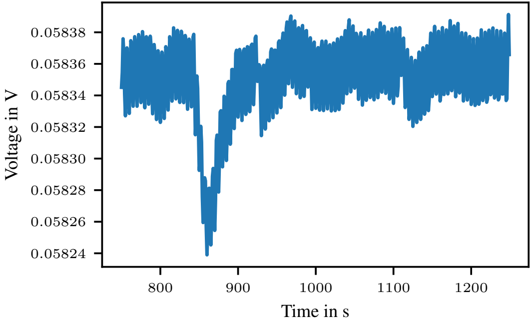
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window



EEG Data

1 Functional Data Analysis

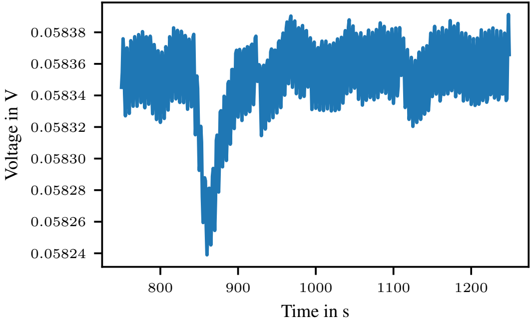
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window
- Methods that analyse EEG data should be shift invariant



EEG Data

1 Functional Data Analysis

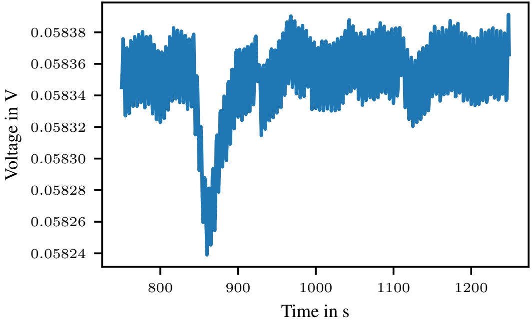
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window
- Methods that analyse EEG data should be shift invariant
- Neural networks can be shift invariant



EEG Data

1 Functional Data Analysis

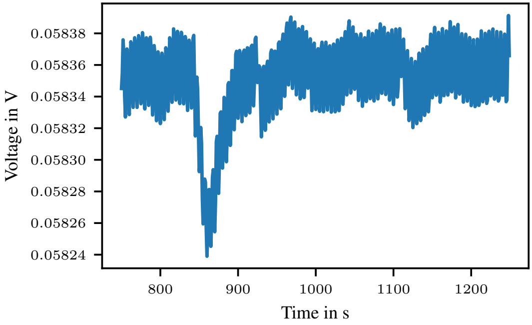
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window
- Methods that analyse EEG data should be shift invariant
- Neural networks can be shift invariant
- In general, neural networks represent data as vectors/matrices



EEG Data

1 Functional Data Analysis

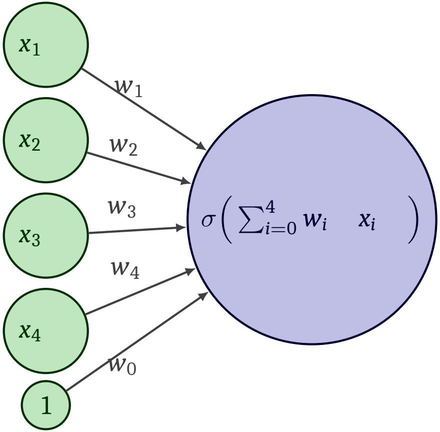
- Electroencephalography (EEG) is a method to measure brain activity
  - Generally (really) noisy
  - Signal of interest is only weak, has a complex pattern and might occur at any time in a given window
- Methods that analyse EEG data should be shift invariant
- Neural networks can be shift invariant
- In general, neural networks represent data as vectors/matrices
- **Combine deep learning with FDA**



Functional Neurons

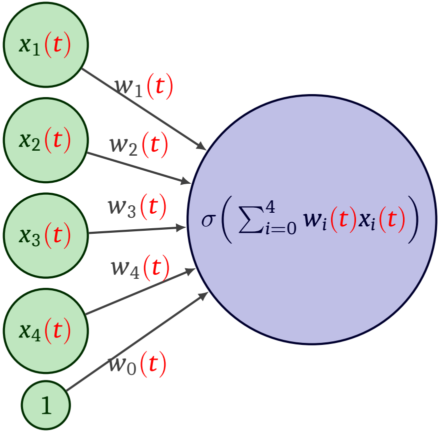
2 Functional Neural Networks

- The output of an artificial neuron is an "activated" linear combination of its inputs



Functional Neurons

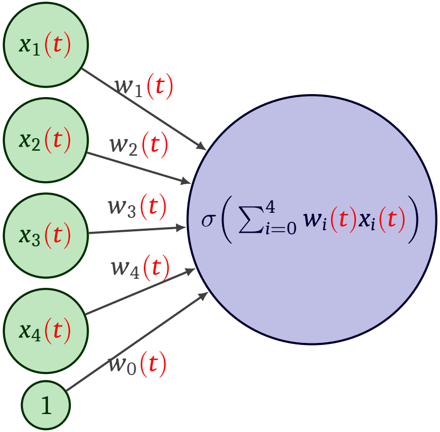
- The output of an artificial functional neuron is an "activated" linear combination of its functional inputs



Functional Neurons

2 Functional Neural Networks

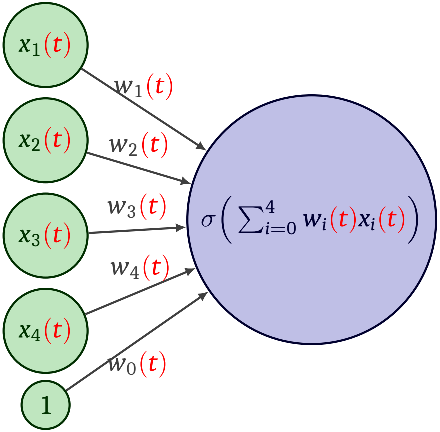
- The output of an artificial functional neuron is an "activated" linear combination of its functional inputs
- Functional neurons reduce the size of the network and the dimension of its input
  - Consider 4 seconds of an 8-channel EEG with sample frequency 250Hz
  - Vector representation: 8000 numbers
  - Functional representation: 8 functions



Functional Neurons

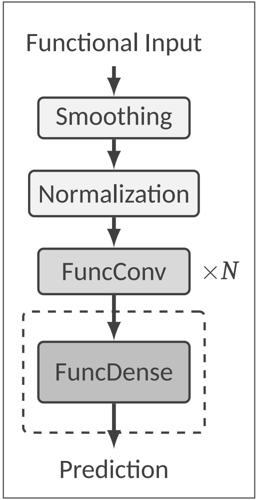
2 Functional Neural Networks

- The output of an artificial functional neuron is an "activated" linear combination of its functional inputs
- Functional neurons reduce the size of the network and the dimension of its input
  - Consider 4 seconds of an 8-channel EEG with sample frequency 250Hz
  - Vector representation: 8000 numbers
  - Functional representation: 8 functions
- Other layers can be extended too, e.g. convolutional or pooling layers



Architecture

- Smoothing: Local linear estimation
- Normalization: Standardizing per channel
- FuncConv: N layers with M filters
  - e.g., $N = 1, M = 20$
- FuncDense: One neuron per class



Empirical Results

2 Functional Neural Networks

- Datasets
  - 2 FDA benchmarks
  - 1 EEG benchmark
  - 2 simulated datasets
- Methods: Neural networks & FDA methods
- Multiple repetitions with random splits
- FNNs achieved state of the art for benchmark datasets

Model	Accuracy	Recall	Precision
Tecator Dataset			
CNN	73.56	71.77	81.32
MLP	85.30	83.80	82.31
FNN(10)	**100.00**	**100.00**	**100.00**
FNN(20)	**100.00**	**100.00**	**100.00**
Berrendero et al.	99.53	-	-
Phoneme Dataset			
CNN	83.73	83.83	85.01
MLP	79.95	80.55	80.05
FNN(10)	**91.53**	91.42	91.56
FNN(20)	**91.53**	**91.52**	**91.72**
Berrendero et al.	81.14	-	-
BCI Competition IV Dataset 2A			
EEGNet	51.81	40.21	44.75
FNN(20)	55.43	44.98	48.31
FNN(40)	**56.29**	**45.95**	**49.46**

Functional Neural Networks

Shift invariant models for functional data with applications to EEG classification

- Poster: https://icml.cc/virtual/2023/poster/25224
- Paper: https://openreview.net/pdf?id=vvcJCbxxbp

Literature

- Berrendero, J. R., A. Cuevas, and J. L. Torrecilla (2016).
  The mrmr variable selection method: a comparative study for functional data.
  Journal of Statistical Computation and Simulation 86(5), 891–907.

- Hastie, T., A. Buja, and R. Tibshirani (1995).
  Penalized discriminant analysis.
  The Annals of Statistics 23(1), 73–102.

- Tangermann, M., K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, et al. (2012).
  Review of the BCI competition IV.
  Frontiers in neuroscience, 55.

- Thodberg, H. H. (2015).
  Tecator meat sample dataset. statlib datasets archive.

Functional Neural Networks

Thank You!