

Tutorium 10

☰ Tags

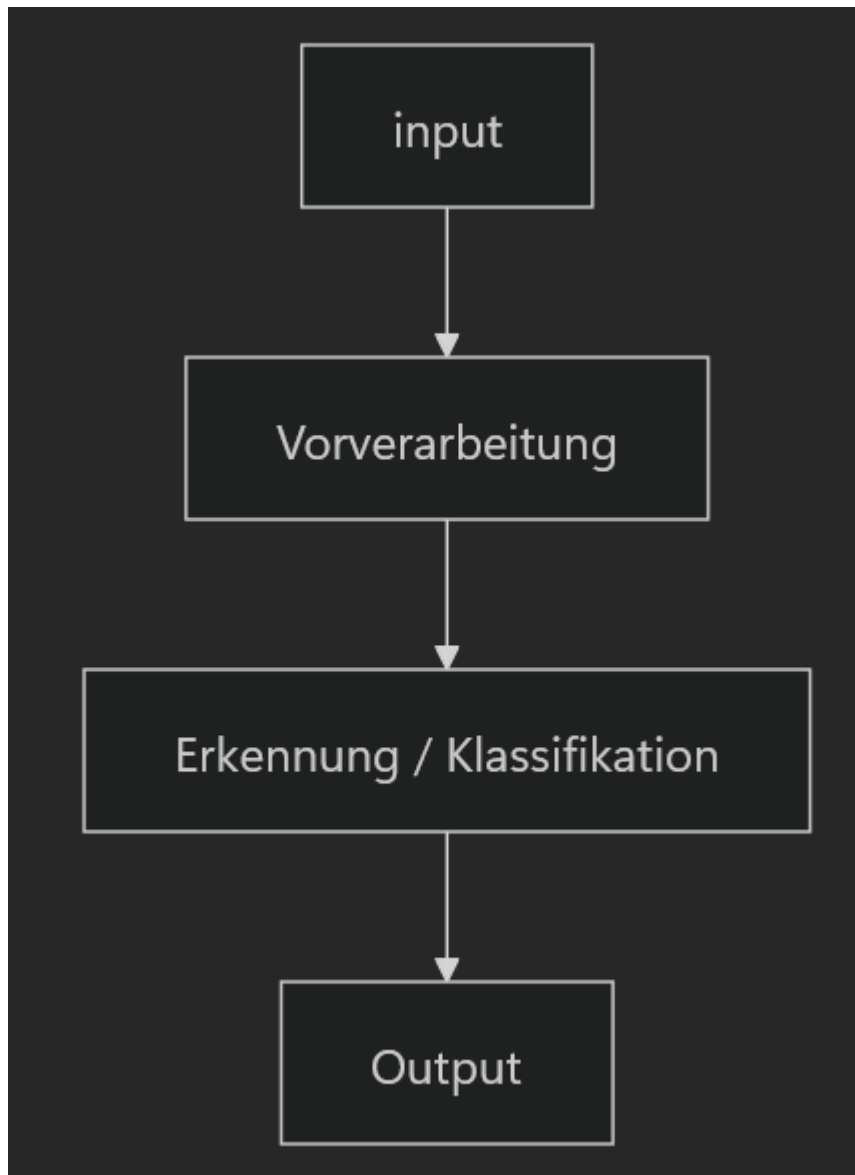
Was ist das Problem, wenn man beim Airwriting nur nach Sequenzen filtert?

- Zögern wird als Wortlücke identifiziert
- Nur unterscheiden zwischen Bewegung und Nicht-Bewegung: Jede Bewegung würde als Wort identifiziert werden, also auch Bewegungen, die gar nicht schreiben
 - 3-5 Hz Bewegungen oft schreiben

Blockschaltbild

- Aufnahmephase ist Inputblock, aufgenommen durch IMU
- Spottingphase gehört tendenziell in die Vorarbeitungsphase
- Erkennungsphase auch ein Block
- Ausgabephase

```
graph TD
    input --> Vorverarbeitung --> A[Erkennung / Klassifikation] --> Output
```



→ Kommt ziemlich sicher in der Klausur, wird nächste Woche weiter gemacht.

Auswahl von Merkmalen

Welche Merkmale sind gute Merkmale?

- objektiv

- diskriminierend (unterscheidet verschiedene Typen gut)
- sollen für gute Klassifizierung sorgen

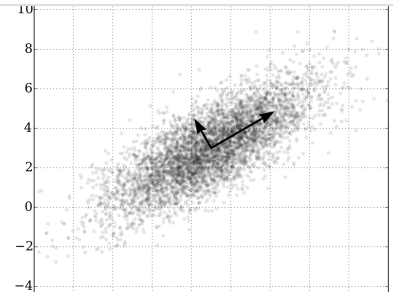
Welche Verfahren eignen sich zur Auswahl von Merkmalen?

- PCA (Primary Component Analysis)
 - reduziert Merkmale

Principal component analysis

Principal component analysis (PCA) is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data

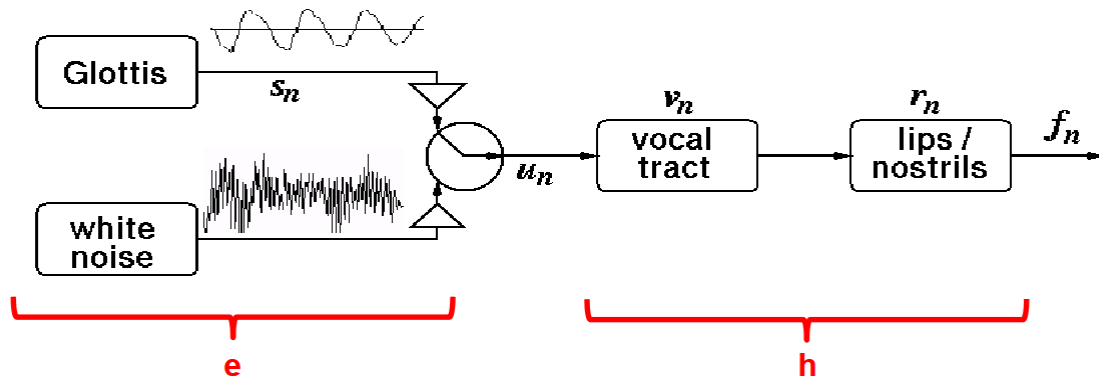
https://en.wikipedia.org/wiki/Principal_component_analysis



- Wenn man die einzelnen Labels der Punkte kennt: LDA (Linear Discriminant Analysis)
- Feature Selection
 - Unsupervised:
 - Remove strongly correlated feature (z.B. schwimmt im Wasser bei Fischen, alle Fische schwimmen im Wasser)
 - Remove zero-variance features
 - Supervised:
 - Wrapper-Ansatz: Forward, Backward, recursive Selection
 - Filter: Pearson R, F-Score, Mutual information
 - Embedded: autoencoder, LASSO (least absolute shrinkage and selection operator)

Ein Beispiel: Mel Frequency Cepstral Coefficients (MFCCs) für Sprache

- Darstellung wurde über 20 Jahren entwickelt, wirkt sehr intuitiv, war aber ein schwerer Weg



- Die **Anregungsfunktion e** ist die Stimmbandanregung und/oder weißes Rauschen
- Der **Filter h** ist die Modulation dieser Anregung durch den Vokaltrakt
- Die resultierende Modulation f ergibt sich aus der Faltung der Anregungsfunktion e mit dem Filter h :

$$f = e * h$$

- Glottis: Grundschiwingung, entsteht vor allem bei Vokalen, wie zum Beispiel dem aaa-Sound
- white noise entsteht durch Geräusche, bei denen diese Grundschiwingung nicht genutzt wird, wie z.B. dem sch-Sound
 - sch und scharfes s hat weniger Amplitude, weil der Mund nicht so geöffnet ist beim Aussprechen wie z.B. beim aaa-Sound
 - gibt aber auch Sounds die beides haben (u_n), wie zum Beispiel das summende s
- Zungenposition, Lippenformation etc. beeinflusst die Sprache auch noch

- Quelle-Filter-Modell: $f = e * h$

- Im Frequenzbereich:

$$F(f) = F(e) \cdot F(h)$$

$$\log F(f) = \log F(e) + \log F(h)$$

$$F^{-1}(\log F(f)) = F^{-1}(\log F(e)) + F^{-1}(\log F(h))$$

- Damit kann man Sprache entfalten in die **Summe** zweier Anteile:
 - Die niederfrequente periodische Anregungsfunktion e
 - Die Filterung des Vokaltrakts (Formanten) h
- Die Koeffizienten dieser Transformation nennt man **Cepstralkoeffizienten** (engl. **cepstral coefficients**) oder einfach **Cepstrum**
- Das Cepstrum ist die *inverse Fourier Transformation des Logarithmus der Magnitude des Spektrums*
- Anregung und Filter werden also kombiniert auf unterschiedliche Weise
 - Gefaltet im Zeitbereich
 - Multipliziert im Frequenzbereich
 - Addiert im Ceptralbereich

Faltung:

Die **Faltung**, auch **Konvolution** (von lateinisch *convolvere* „zusammenrollen“), beschreibt in der Analysis einen mathematischen Operator, der für zwei Funktionen f und g eine dritte Funktion $f * g$ liefert.

[https://de.wikipedia.org/wiki/Faltung_\(Mathematik\)](https://de.wikipedia.org/wiki/Faltung_(Mathematik))

Faltung noch einmal anschauen

Berechnung der MFCC

