4. Machine Learning and Binaural Hearing

1. Binaural Features
2. **Mapping Sounds onto Their Directions**
3. Collecting Training Data
4. The Binaural Manifold
5. Localization with a Look-up Table
6. Linear Regression
7. Piecewise Linear Regression
8. Predicting the Direction of Speech
9. Principles of Sound Separation
10. Separation & Localization Method

Binaural Features and Source Direction

- Remember the link between IPD (interaural phase difference) and TDOA (time difference of arrival).
- This link is a direct consequence of the time-shift theorem (week #2).
- Also, there is a link between TDOA and source direction in case of direct propagation.
- In the presence of filtering effects, such as the HRTF (head-related transfer function) **there is NO explicit relationship**.

ILPD and Sound-Source Direction

- Let $\boldsymbol{y} \in \mathbb{R}^D$ be an ILPD vector observed with a binaural robot head.
- $D = 3F$, or the concatenation of ILD (of size F) and IPD (of size 2F), for example $F = 512$ frequencies.
- Let $\boldsymbol{x} \in \mathbb{R}^2$ be the direction (azimuth and elevation) of a sound source.

Acoustic to Direction Mapping

- We seek an **explicit** representation of the mapping:

$$\text{ILPD vector} \Longrightarrow \text{source direction}$$

- Or a function f that maps **high-dimensional** ILPD vectors onto **low-dimensional** directions:

$$\boldsymbol{x} = f(\boldsymbol{y}), \quad f : \mathbb{R}^D \longmapsto \mathbb{R}^2$$

Learning an Acoustic-to-Direction Mapping (I)

Unsupervised learning: Both f and $\boldsymbol{x}$ are unknown.

1. Sample the high-dimensional space, $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n, \dots, \boldsymbol{y}_N$;
2. Extract a low-dimensional manifold from the high-dimensional sample space;

- f in $\boldsymbol{x} = f(\boldsymbol{y})$ is a **linear** or **non-linear** projection of $\mathbb{R}^D$ onto $\mathbb{R}^2$.
- This is referred to as:
- **dimension reduction** (PCA) or
- **manifold learning** (LE, LTSA).

Learning an Acoustic-to-Direction Mapping (II)

Supervised learning: Only f is unknown.

1. Sample the high-dimensional space, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n, \ldots, \boldsymbol{y}_N$;
2. Observe sound directions, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \ldots, \boldsymbol{x}_N$;
3. Form an input-output set of training samples, $(\boldsymbol{y}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{x}_n), \ldots, (\boldsymbol{y}_N, \boldsymbol{y}_N)$

- Estimate f from:

$$\begin{aligned} \boldsymbol{x}_1 &= f(\boldsymbol{y}_1), \\ &\vdots \\ \boldsymbol{x}_n &= f(\boldsymbol{y}_n), \\ &\vdots \\ \boldsymbol{x}_N &= f(\boldsymbol{y}_N). \end{aligned}$$

- This is referred to as **regression**.

Regression

- **Linear regression**, a projection of $\mathbb{R}^D$ onto $\mathbb{R}^2$:

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{y} + \boldsymbol{b}, \quad \mathbf{A} \in \mathbb{R}^{2\times D}, \boldsymbol{b} \in \mathbb{R}^2$$

- **Piecewise-linear regression** (there are K possible projections):

$$\boldsymbol{x} = \sum_{k=1}^{K} \mathbb{I}(z = k)(\mathbf{A}_k \boldsymbol{y} + \boldsymbol{b}_k)$$

- $\mathbb{I}(z)$ is called an **indicator function**, that selects the k-th affine transformation $\mathbf{A}_k, \boldsymbol{b}_k$.
- We will built a binaural localization method based on piecewise-linear regression.

Session Summary

- Link between binaural features and sound localization
- Learning an acoustic-to-direction mapping
- Unsupervised learning
- Supervised learning
- Regression