Cognitive Neuroscience II

Prof. Dr. Andreas Wendemuth

Lehrstuhl Kognitive Systeme

Institut für Elektronik, Signalverarbeitung und Kommunikationstechnik

Fakultät für Elektrotechnik und Informationstechnik
Otto-von-Guericke Universität Magdeburg

http://iesk.et.uni-magdeburg.de/ko/

Lecture 16
Representational Learning

- Prior / Posterior distributions
- Probability densities as generative models

Representations

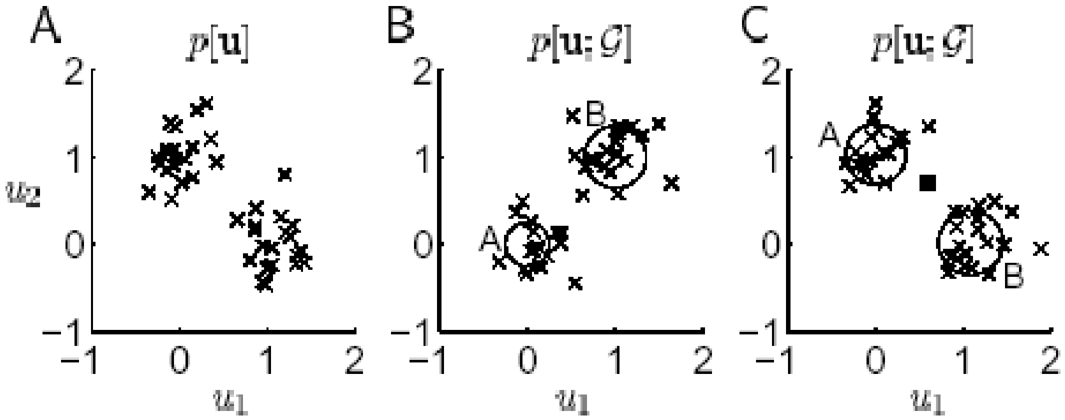
- Sensory representations serve as prime input to neurons (e.g. pixels)
- A re-representation is needed for data compression (i.e. parameters of a normal distribution)
- We ask:
  - What goals are served by particular representations?
  - How are representations developed on the basis of the input statistics?

Generative / Recognition Model

- Prime input (e.g. pixels) is highly structured by „causes“ (e.g. solid objects)
- *Capture* this structure by „generative model“ (e.g. normal distribution)
- *Identify* structure (e.g. recognize objects) by „recognition model“ (e.g. classification)

Generative Model: Gaussian Mixtures

- Let N(**x**|**m**,**S**) be a normal distribution of **x** with mean **m** and Covariance Matrix **S**.
- Then $P(\mathbf{x} \mid \Theta) = \sum_{k=1}^{K} c_k N(\mathbf{x} \mid \mathbf{m}_k, \mathbf{S}_k)$ with $1 = \sum_{k=1}^{K} c_k$ is a Gaussian mixture model. Looks like



Gaussian Mixtures (2)

- In neuronal terms:
  Without any further knowledge, neuron j produces a spike x with *prior probability* or prior distribution over causes $c_j = P(j)$.
- The *generative distribution* is the probability $P(\mathbf{x}|\mathbf{j}, \mathbf{m}_\mathbf{j}, \mathbf{S}_\mathbf{j})$ that, given neuron j and parameters $\mathbf{m}_\mathbf{j}, \mathbf{S}_\mathbf{j}$, the spike was produced by neuron j.

Gaussian Mixtures (3)

- Hence the Gaussian Mixture model

$$P(\mathbf{x} \mid \mathbf{\Theta}) = \sum_{k=1}^{K} c_k N(\mathbf{x} \mid \mathbf{m}_k, \mathbf{S}_k)$$

$$= \sum_{k=1}^{K} p(k) P(\mathbf{x} \mid k, \mathbf{m}_k, \mathbf{S}_k)$$

is a *sum over causes*, in statistics a marginal distribution with parameters $\mathbf{\Theta}$.

Gaussian mixtures (4)

- This can be written as a sum of *joint distributions* (joint: causes and generations)

$$P(\mathbf{x},k\,|\,\boldsymbol{\Theta}_k) = p(k)P(\mathbf{x}\,|\,k,\mathbf{m}_k,\mathbf{S}_k)$$

as

$$P(\mathbf{x}\,|\,\boldsymbol{\Theta}) = \sum_{k=1}^{K} P(\mathbf{x},k\,|\,\boldsymbol{\Theta}_k)$$

Recognition/decision

- We would like to know with which probability P(k|x) the activity x was produced by neuron k. Apply Bayes Rule:

  $$P(\mathbf{x},k) = P(k)P(\mathbf{x}\,|\,k) = P(k\,|\,\mathbf{x})P(\mathbf{x})$$

  hence the *posterior probabilities* are

  $$P(k\,|\,\mathbf{x}) = \frac{P(k)P(\mathbf{x}\,|\,k)}{P(\mathbf{x})} = \frac{P(\mathbf{x},k)}{P(\mathbf{x})}$$

  and the maximum is to be selected.

Decisions

- In order to make a decision, the denominator P(**x**) is irrelevant. As an *indicator function*, we can use the likelihood (numerator) with $P(\mathbf{x},k) = p(k)P(\mathbf{x}\,|\,k,\mathbf{m}_k,\mathbf{S}_k) = c_k N(\mathbf{x}\,|\,\mathbf{m}_k,\mathbf{S}_k)$ and select the maximum. So Bayes‘ decision is equivalent to Maximum Likelihood decision.

Ex 6/1

- Given are the 14 points of 2-dim. data (-2,4),(-1,-1),(-1,1),(0,0),(1,-1),(1,0),(1,1), (1,2),(2,1),(2,2),(2,3), (3,2),(3,4),(6,0)
- Draw the data. By inspection, assign 2 mixture densities to the data. What are the means m_k and relative weights c_k of the 2 components?
- With your values of m_k and c_k, compute the covariance matrices.
- Does a coordinate transformation exist (try inspection first!, otherwise calculate) which makes the covariance matrices diagonal? Without computing covariance matrices: what is the geometrical justification for such a transformation?

Ex 6/2

- What is the probability (formula only) that a data point is generated anywhere in (5±1, 0±1)?
- Approximate that probability, using the Gaussian Mixture probability density at (5,0), and the value of the inspected area.
- What are the probabilities (formula only) that a data point anywhere in (5±1, 0±1) is generated by mixture component 1 (2)?
- Approximate these probabilities, acting as above.
- What is the likeliest cause for that data point?

Clustering

- Gaussian Mixtures effectively provide a clustering of data into the mixture components.
- To fix the parameters $\mathbf{m}_j$, $\mathbf{S}_j$, we need some training algorithm.
- Here, we look at a simplified version called the K-means algorithm

K-means

- Want to find *regions* (not: probability distributions) for K clusters in an unsupervised fashion. <u>Algorithm:</u>
- Randomly arrange K means about global mean
- Repeat (until fluctuations become low):
  – Define regions by smallest distance to means
  – Compute new means from all points in region

Ex 7

- Given the data points in ex 6/1, perform a K-means clustering with 2 classes. You can do this by drawing and/or by computing.
- Does that clustering give the same most likely cause as obtained in ex 6/2?
- Explain the difference between discriminative clustering (k-means) and maximum likelihood clustering (using a pdf).

Resumé of lecture 16

- Raw data, if produced by causes, needs re-representation to be compactly and meaningfully captured in a generative model.
- Gaussian Mixture densities can serve as good pdf‘s for generative models.
- If classification only is required, space partitioning methods such as k-means are useful.