WS 2015/2016

Machine Learning I – Probabilistic Unsupervised Learning

Jörg Lücke (lecturer), Maryam Sadreddini (TA)

Planned Lecture Content

Lecture 1 – Introduction to Machine Learning and basic probability
- introductory talk on probabilistic modeling
- probability types, marginalization, Bayes' rule

Lecture 2 – Expectations, Densities and Sampling
- expectation values, probability densities, data samples
- approximating expectation values by sampling

Lecture 3 – Density Estimation and Classification
- samples and density parameters (Gaussian case)
- optimal classification
- data likelihood

Lecture 4 – The EM Algorithm (and some mixture models)
- problems: likelihood and mixture model
- excursion: Jensen's inequality
- free-energy, entropy, Kullback-Leibler divergence
- the EM algorithm

Lecture 5 – The Mixture of Gaussians Model
- excursion: constraint optimization
- parameter update rules, learning algorithm
- local optima and a theory bug

Lecture 6 – K-means, Single-Cause and Multiple-Cause Models
- k-means
- mixture models as single-cause models
- data points as images, and images as data points
- introduction to multiple-cause models
- example: binary sparse coding
- notes on the generalized EM algorithm

Lecture 7 – A Learning Algorithm for Binary Sparse Coding and PCA

- EM learning for binary sparse coding
- computational intractability
- numerical examples

- Principal Components Analysis (PCA)
- definition of the problem
- the deterministic solution in one dimensions

Lecture 8 –

- Deterministic and Probabilistic PCA
- deterministic PCA solution for k dimensions
- discussion and examples

- PCA as a probabilistic generative model (p-PCA)
- continuous hidden variables
- EM learning for p-PCA
- example implementation (show matlab code)
- discussion of the probabilistic solution
    - noise and variance
    - factor analysis
    - computational advantages
    - dimensionality reduction, compression
    - equivalence to mean-squared error

Lecture 9 – Sparse Coding and Independent Components Analysis (ICA)

- the sparse coding generative model
- EM for standard sparse coding
- required approximation (MAP)
- example implementation (show matlab code)
- ICA as noiseless limit
- example applications
- models of neural processing

Lecture 10 – Sparse Coding and Approximate Inference and Learning

- what can we do if we do not have an analytical solution
- when it wasn't so perfect: MAP, sampling (example Sparse Coding)
- approximation schemes more systematically:
    - MAP approximation (standard for sparse coding)
    - Laplace approximation
    - variational approximation
    - stochastic approximations (sampling)

X-MAS BREAK

Lecture 11 Hidden Markov Models (1/3)
– Markov and Hidden Markov Models
- Markov Models, transition matrices etc
- maximum likelihood solution for Markov Models
- Hidden Markov Models, motivation
- graphical model, generative model, joint probability

Lecture 12 – Hidden Markov Models (2/3) – Inference
- types of inference, filtering, smoothing, best hidden series
- filtered marginals: the Forwards Algorithm
- smoothed marginals: the Forwards-Backwards Algorithm

Lecture 13 – Hidden Markov Models (3/3) – Inference and Learning
- learning of model parameters
(EM for HMMs or the Baum-Welch Algorithm)

Lecture 14 – Overview, Conclusions, Discussions and Outlook
HMMs continued
- the most likely sequence of hidden states
(Viterbi Algorithm, just mentioned)
Summary of HMMs
Summary: Models and Algorithms – Probabilistic/Non-Probabilistic
- relation between the treated models
- differences between probabilistic and non-probabilistic approaches
(unified view within a probabilistic framework)
- models that have not been treated (outlook Machine Learning II)- -

What are the challenges?
- deterministic and stochastic approximations
- approximations as key to new models
- some hidden secrets
- examples of new models

- outlook Machine Learning II

Concluding Remarks

THANKS!