Interconnection of Restricted Boltzmann Machine method with statistical physics and its implementation in the processing of spectroscopic data

Author: Bc. Jakub Vrábel

Supervisor: Ing. Pavel Pořízka, Ph.D.

Content

- Motivation
- Machine Learning + Artificial Neural Networks
- Interconnection with statistical mechanics
- Restricted Boltzmann Machine (RBM)

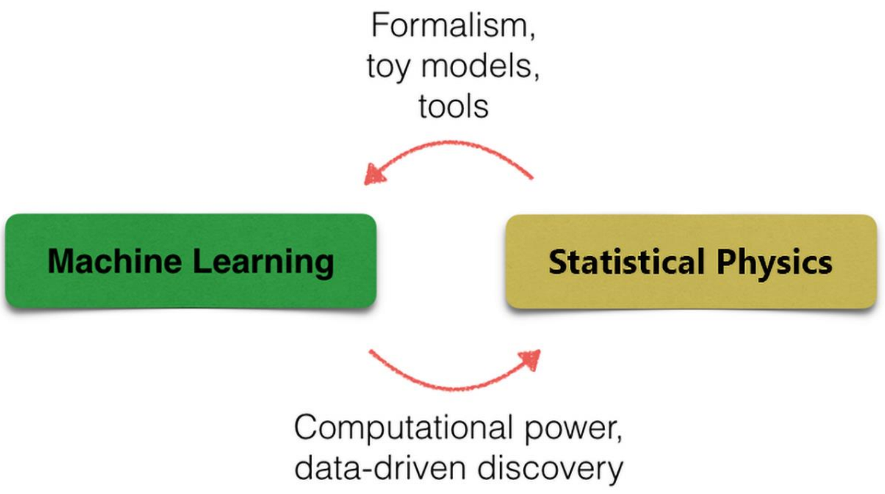
- Spectroscopic data
- Dimension reduction by RBM
- Further plans

Motivation

- Machine Learning
  - Data-driven world
  - Parametric models
  - Artificial Neural Networks success story

- Appl. to scientific data
  - Spectra classification
  - Phase transition detection
  - Many more

- Statistical mechanics
  - Collective behavior
  - Atoms, spins, bits, …



M. Koch-Janusz (2018, edited)

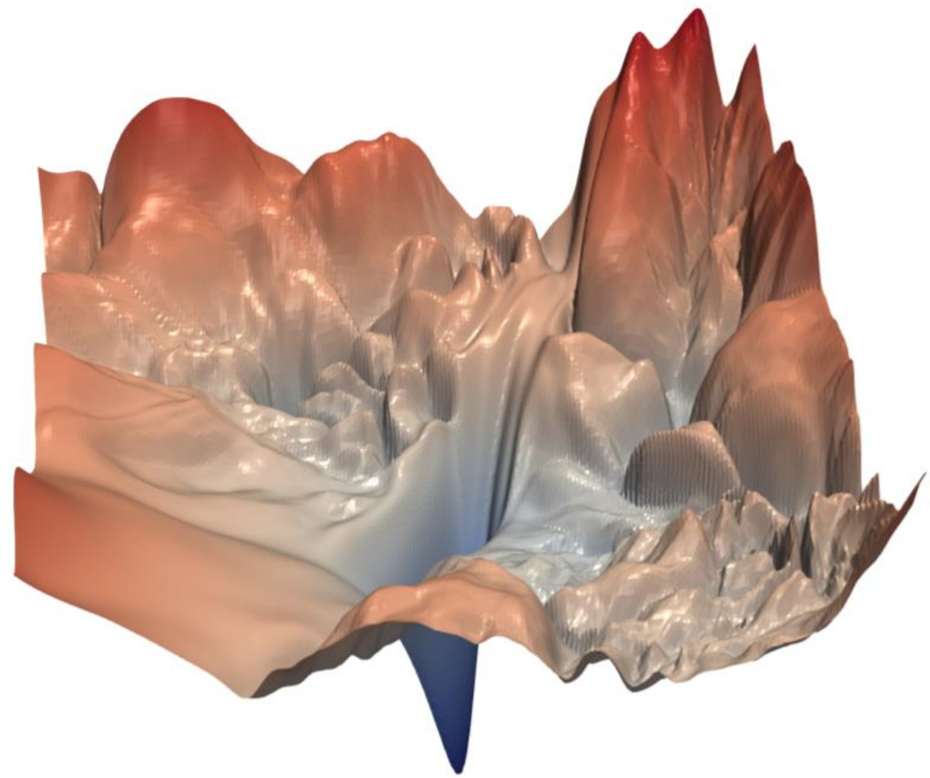
Machine Learning

- Parametric model $g(\boldsymbol{w})$
- Cost (error) function

$$\mathcal{C}(\boldsymbol{X}, g(\boldsymbol{w}))$$

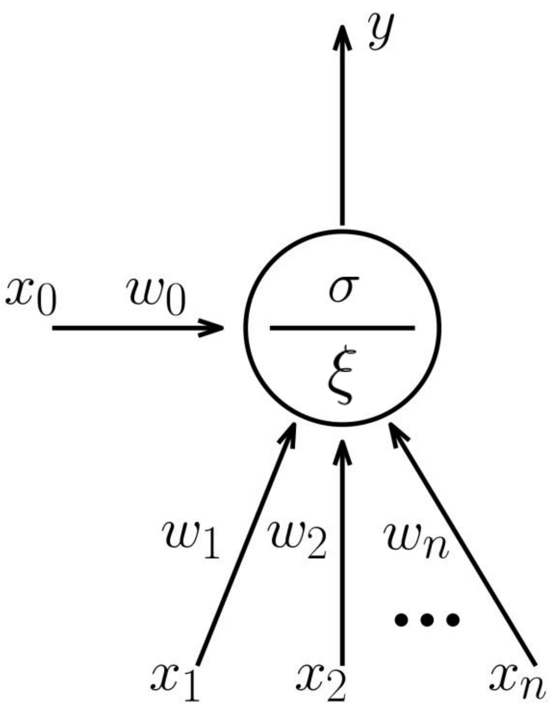
data model

- Learning process
  - Adjusting parameters
  - Gradient descent
- Supervised
- Unsupervised
  - Search for regularities or patterns in data

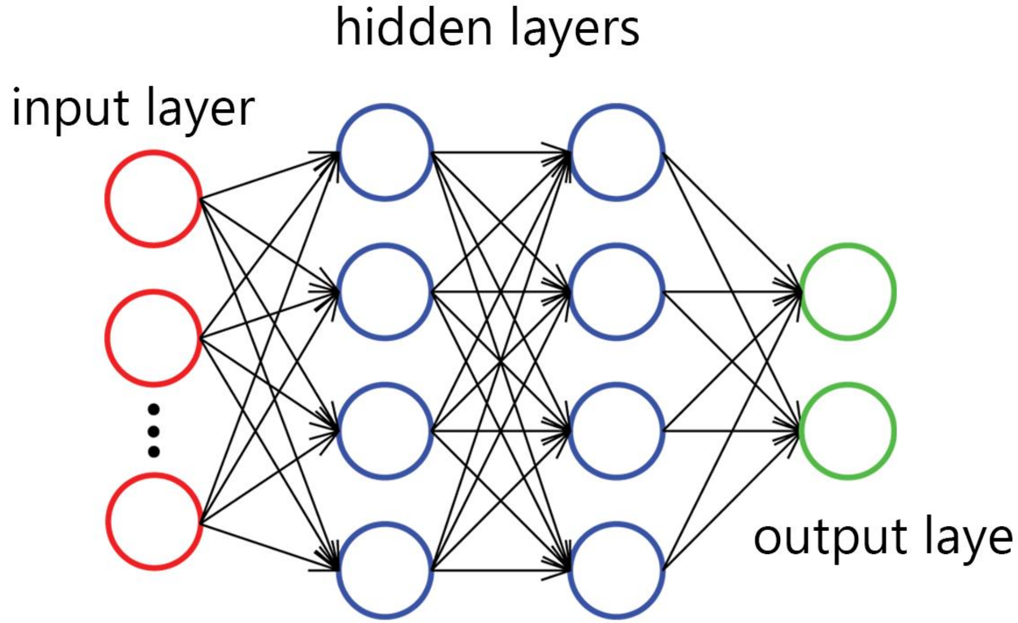


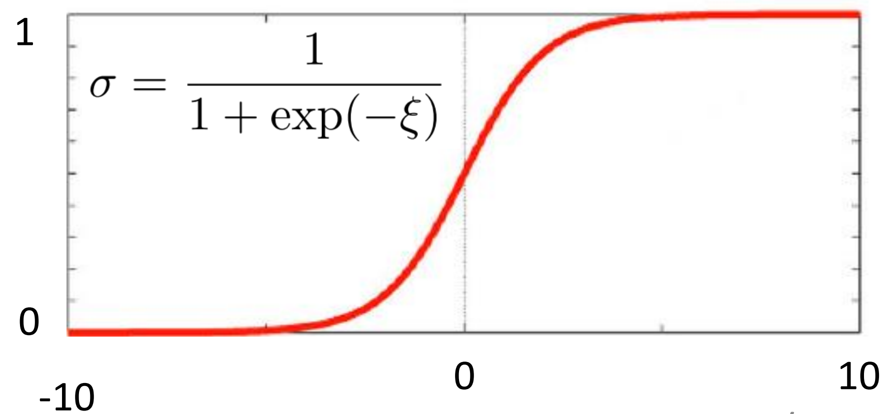
Error landscape (T. Goldstein 2018)

Artificial Neural Networks (ANN)



$$\xi = w_0 + \sum \underbrace{w_i}_{\text{weight (connections)}} \underbrace{x_i}_{\text{input}}$$





Statistical mechanics - introduction

Shannon entropy (1948)

$$S = -\sum_i p_i \log p_i$$

Boltzmann distribution

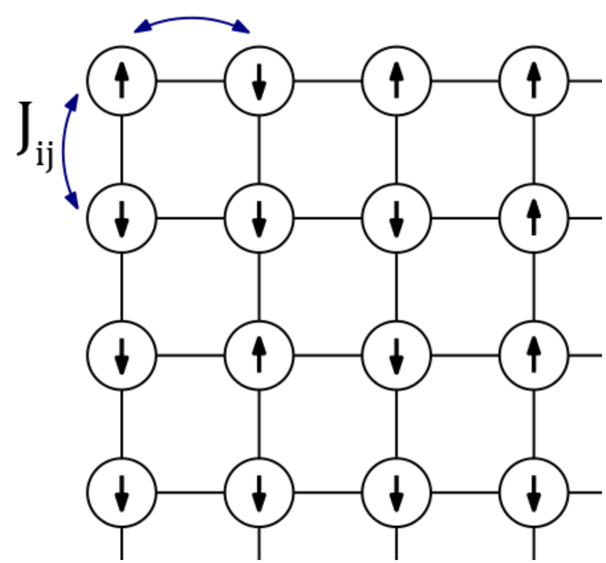
$$p_i = \frac{\exp(-\beta E_i)}{Z}$$

Max. entropy principle

Partition function

$$Z = \sum_i \exp(-\beta E_i)$$

2D Ising model



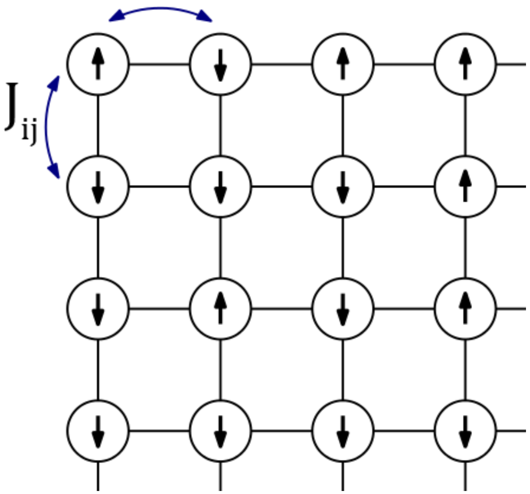
Energy of specific configuration

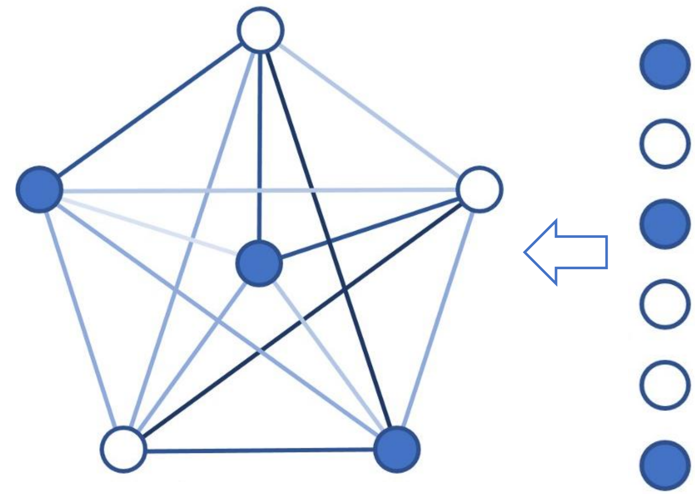
$$E[s] = -\frac{1}{2}\sum_{i\sim j} J_{ij} s_i s_j - \sum_i h_i s_i$$

2D + external field

- intractable partition function

Neural Networks and Machine Learning





(Picture from XanaduAI website)

Ising model – minimization of free energy

- Partition function?

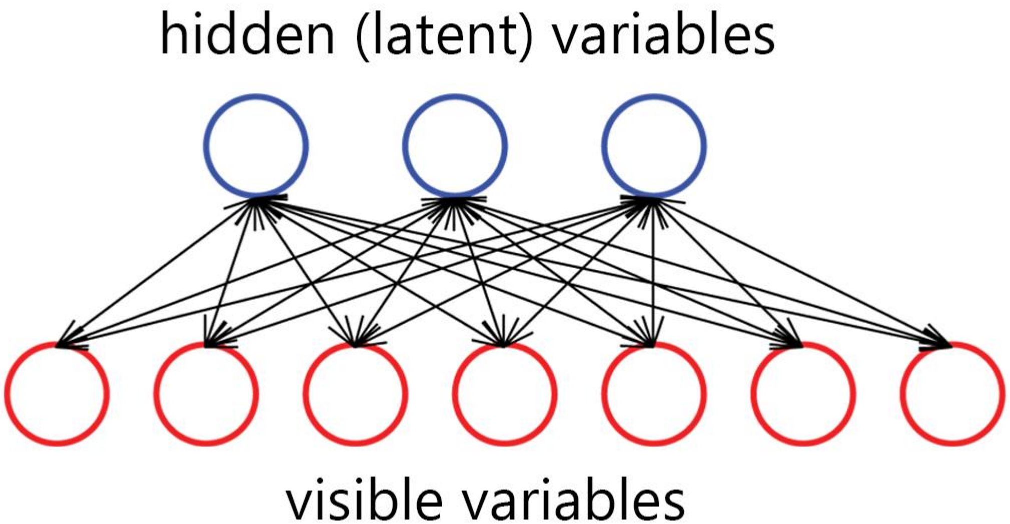
$$E[s] = -\frac{1}{2}\sum_{i\sim j} J_{ij} s_i s_j - \sum_i h_i s_i$$

Hopfield network

- Optimization tasks
- Associative memory (limits)
- Correlations (only 1st order)

Variational free energy min. $$a_m = \beta \left(\sum_n J_{mn}\bar{x} + h_m \right) \qquad \bar{x}_n = \tanh(a_n)$$

Restricted Boltzmann Machine (RBM)



Learning probability distr.:
- Gibbs sampling

Latent (hidden) variables
- Capturing correlations
- Dimension reduction

Generative model

$$E(\mathbf{v},\mathbf{h}) = -\sum_i \underbrace{a_i v_i}_{\text{visible units (data)}} - \sum_\mu \underbrace{b_\mu h_\mu}_{\text{hidden units}} - \sum_{i\mu} \underbrace{w_{i\mu} v_i h_\mu}_{\text{interaction (connections)}}$$

Adjusting parameters to Minimize Kullback–Leibler divergence

$$D_{KL}(Q||P) = \sum_x Q_\theta(x) \ln \frac{Q_\theta(x)}{P(x)}$$

Spectroscopic data

High-dimensionality

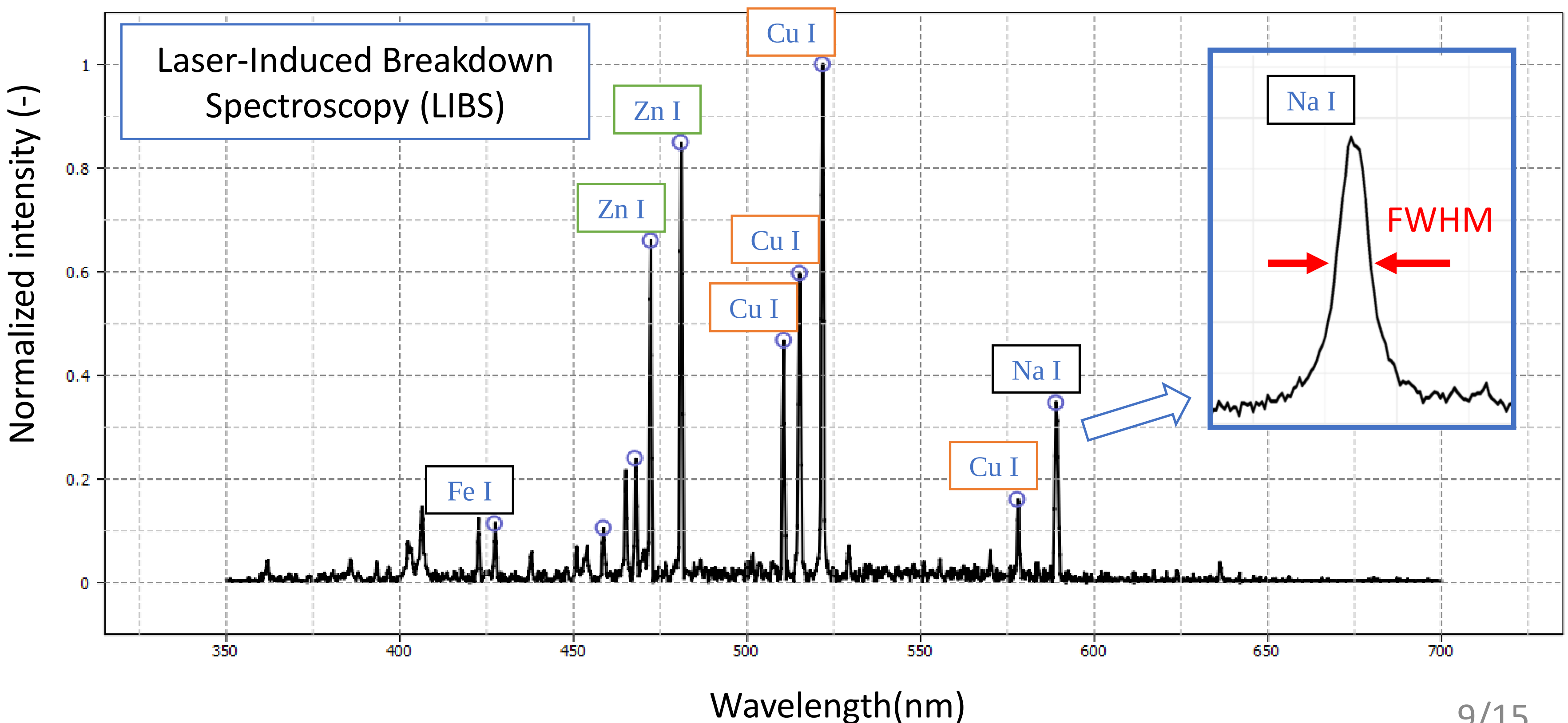
- Spectrograph resolution

Sparsity

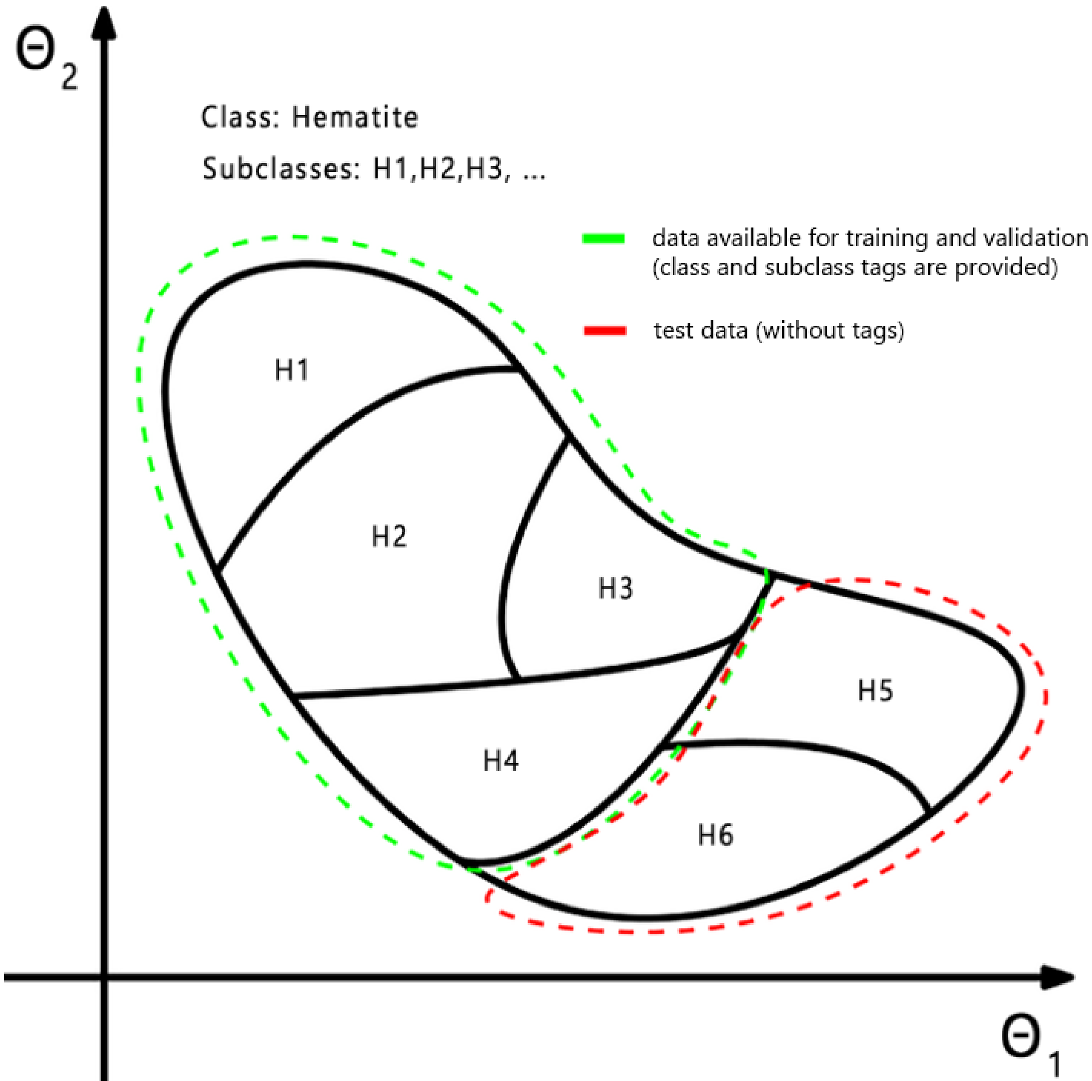
- Useful signal
- Background and noise

Redundancy

- Multiple spectral lines
- 3 values per line



Experiment and data



LIBS measurement

- SciTrace instrument
- CEITEC BUT

Samples

- OREAS certified soils
- Casted to gypsum

Data

- 138 unique samples
- 5000 spectra/sample
- 12 classes (categories)

https://github.com/JVrabel/Diploma_thesis_attachements

Data processing

Programming languages used

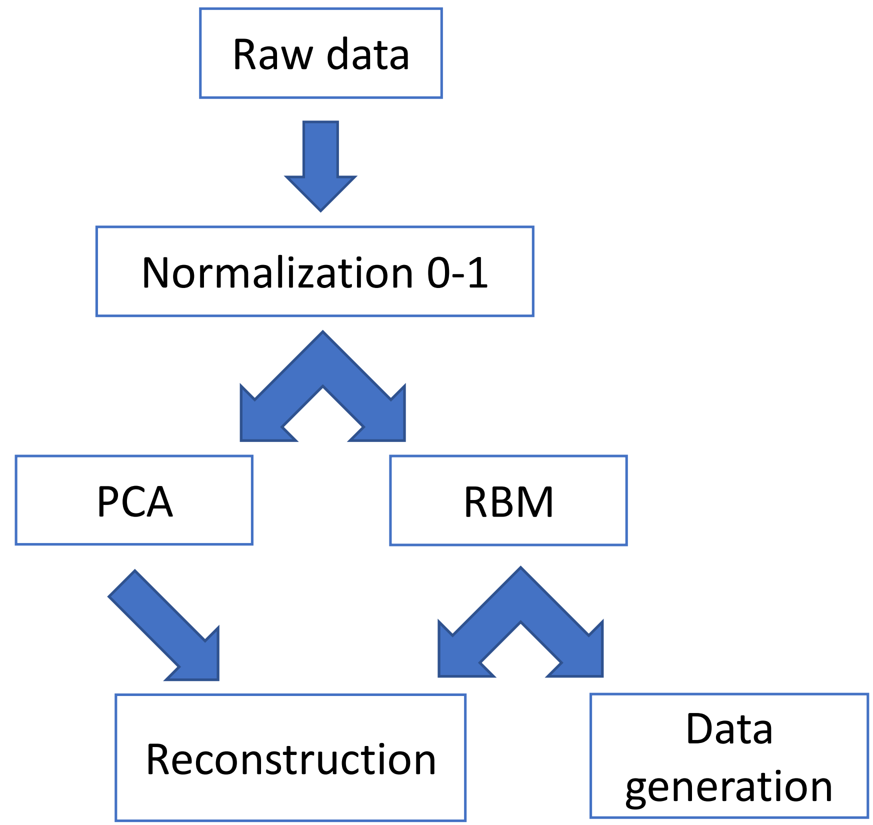
- Python (+TensorFlow)
- R

3 computational scripts written

- 2 – RBM
- 1 – PCA

+ Supplementary scripts

- Visualization
- Preprocessing

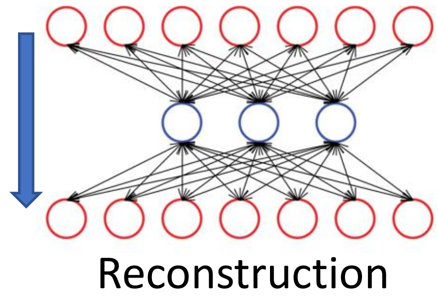


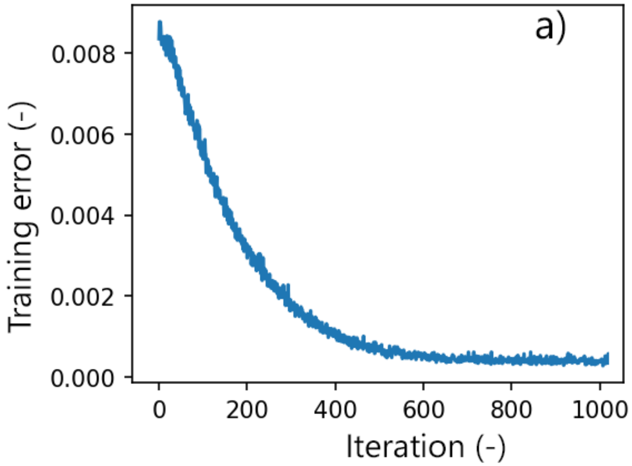
https://github.com/JVrabel/Diploma_thesis_attachements

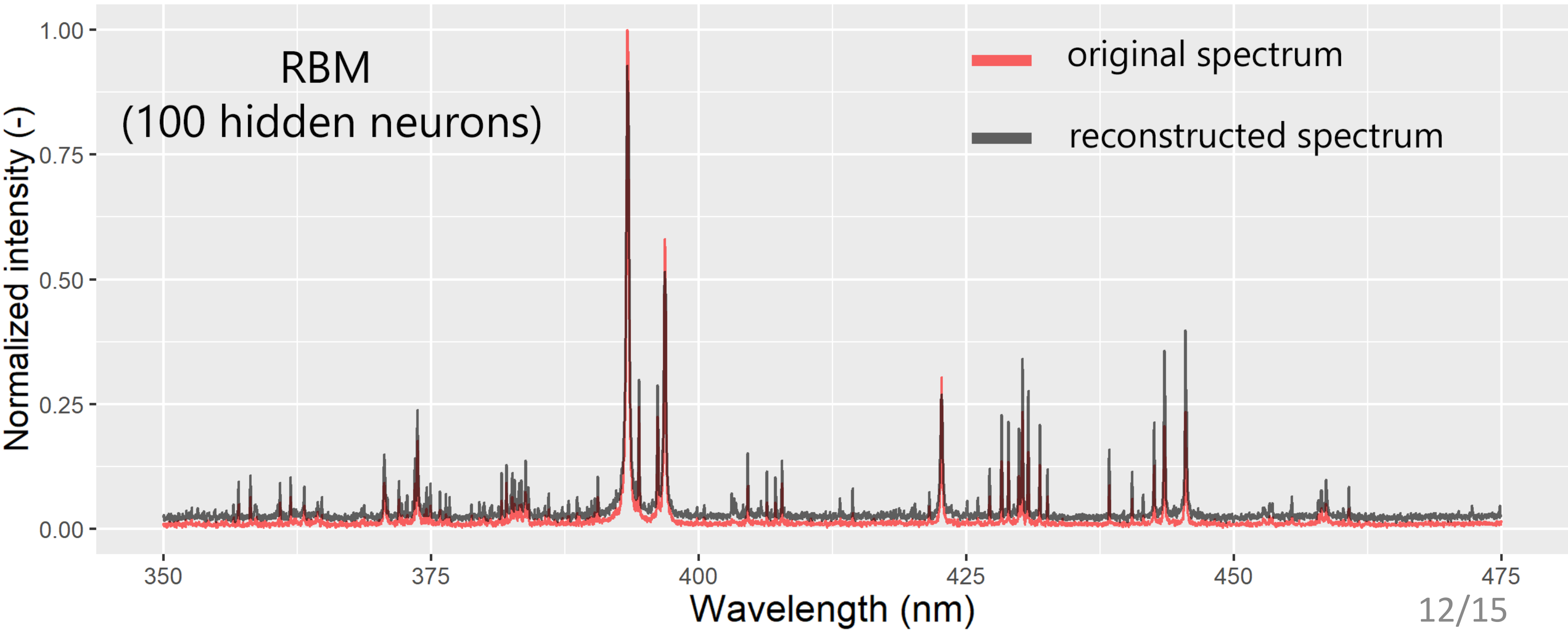
Dimension reduction

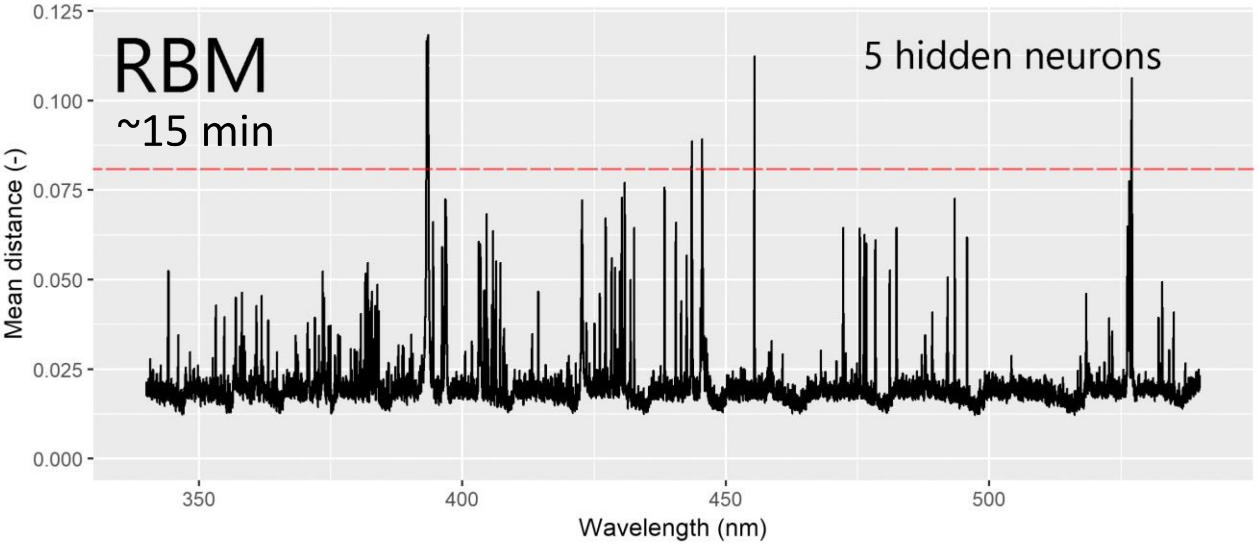
- 30 000 spectra
  - 30 samples
  - 2 classes
- Dimension reduction

  10 000 → 100









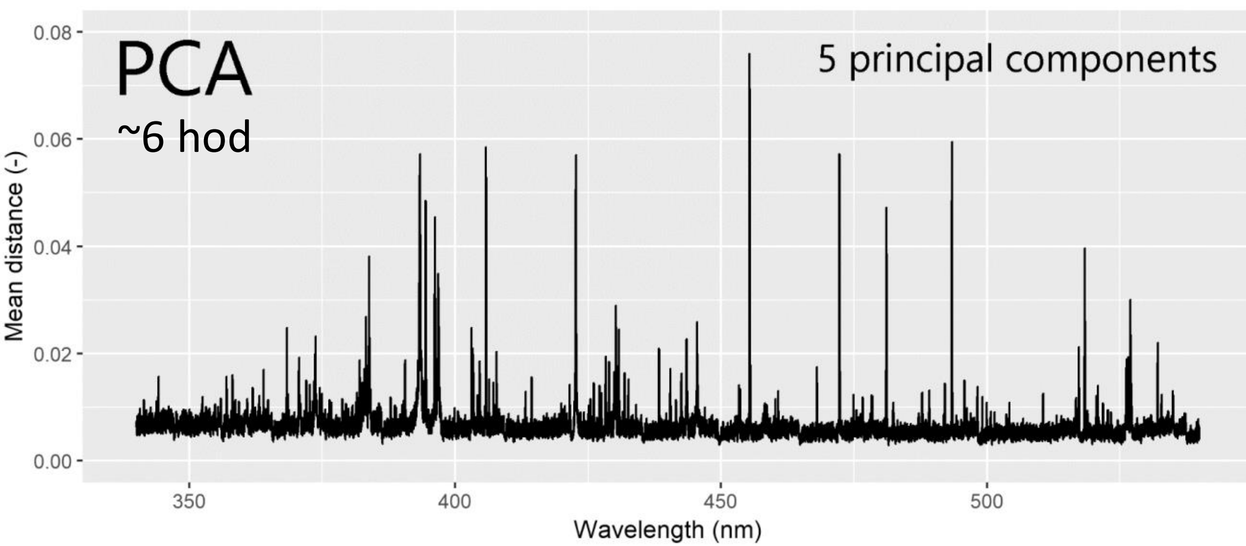
Performance comparison

- 100 samples
- 10 000 spectra in total

Mean distance (L1 norm)

- Computed of 100 representants

	performance	speed	interpretability	extensibility
RBM	8 / 10	9/10 ↑	6/10	8/10 ↑
PCA	9 / 10 ↑	3/10 ↓	10/10 ↑	5/10 ↓

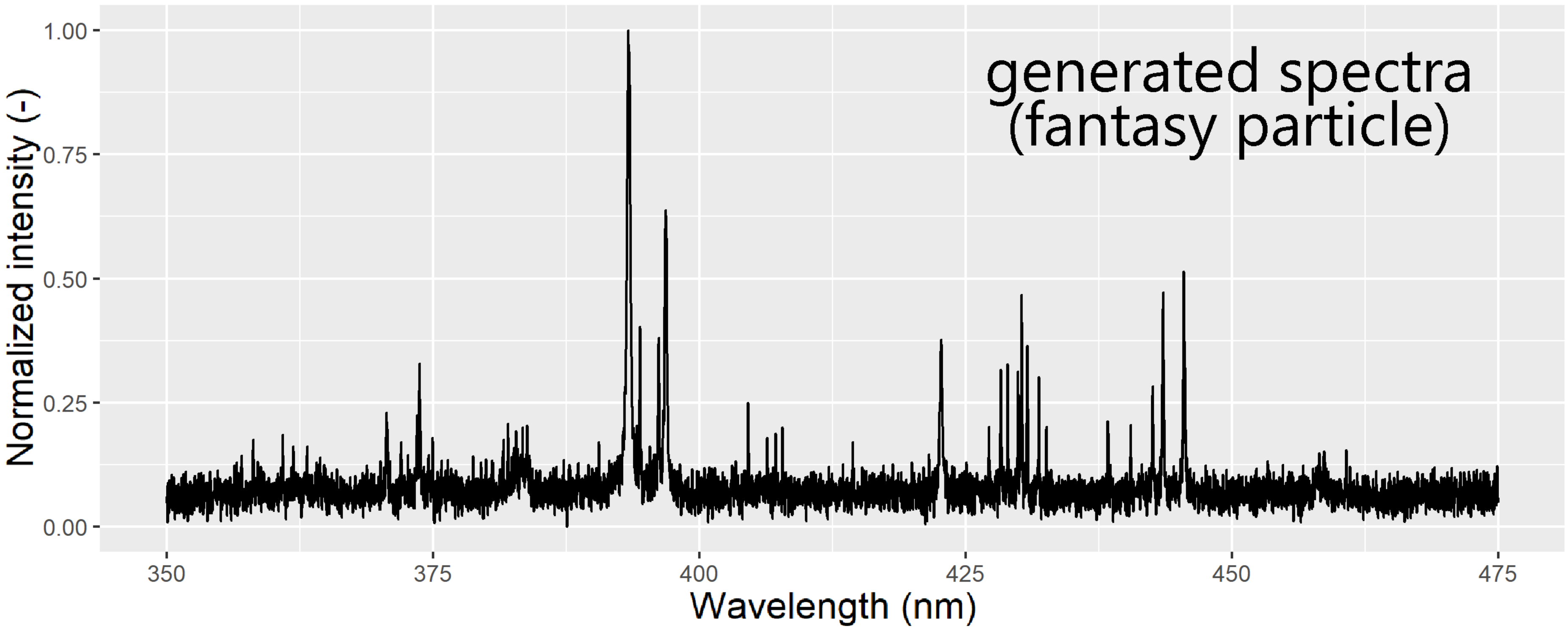


Principal component analysis (PCA)

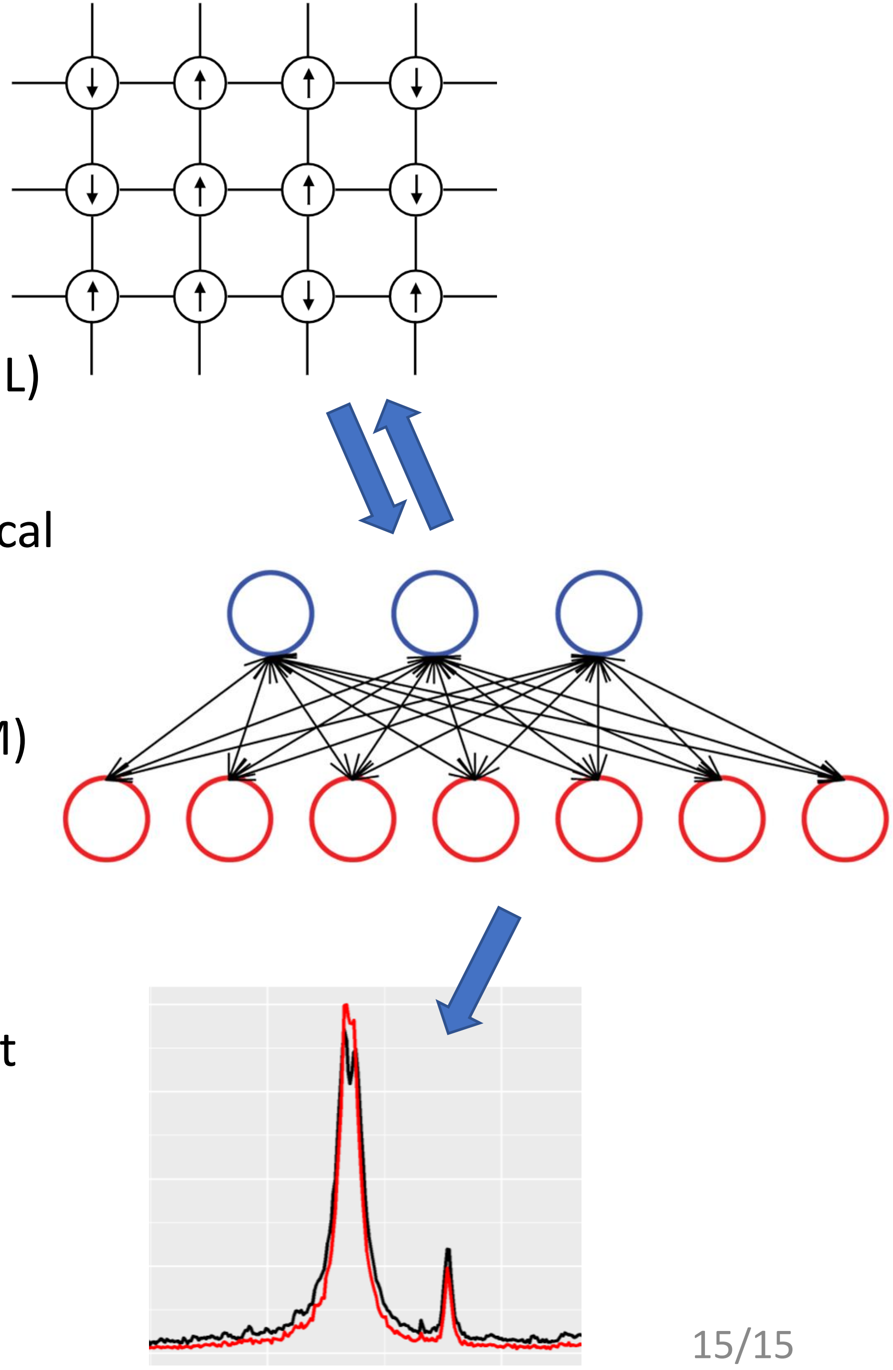
- Linear method
- New variables (basis)

Generovanie nových spektier pomocou RBM

- 30 000 spectra
  (30 samples, 2 classes)
- Dimension reduction
  10 000 → 100



Summary



- Introduction to machine learning (ML)
- Connection between ML and statistical mechanics
- Restricted Boltzmann Machine (RBM)
- Spectroscopic data
- Dimension reduction of huge dataset

Further plans and improvements

1. otázka oponenta

Na str. 30 správně zmiňujete předpokládané gaussovské rozdělení chyb experimentálních dat. Taková data však často vykazují odchylky od normálního rozdělení. Můžete uvést příklady metod strojového učení, kdy tato skutečnost ovlivňuje a kdy neovlivňuje výsledky těchto procedur?

Je pravdou, že v metóde LIBS sa pomerne často vyskytuje rozdelenie extrémnych hodnôt (Extreme Value Distribution) a to v intenzite jednotlivých spektrálnych čiar (viz napr. A. Michel, 2007). Problémom takéhoto rozdelenia dát, je neexistencia rozptylu alebo strednej hodnoty.

Jednoduché lineárne modely, založené na tradičnej štatistike môžu mať problém s takýmto rozdelením. Naopak, pokročilejšie modely strojového učenia, si v prípade dobrej „zobecnitelnosti“ (generalizability) ľahko poradia aj s týmto typom dát.

V spomenutej publikácii, autor navrhuje použitie metódy maximálnej vierohodnosti (Maximum Likelihood Estimation), pre riešenie problému. Metóda RBM implementuje komplexnejšiu verziu MLE a je teda prirodzene vhodná na toto použitie.

2. otázka oponenta

Jaké jsou hardwarové požadavky na zpracování celých spekter LIBS z echelle typicky 200-1000 nm při 5-100 skrytých neuronech u RBM? Co myslíte velkým počtem dat (str. 56) pro PCA?

Na spracovanie dát bol použitý bežne dostupný osobný PC s priemerným procesorom a 32 GB RAM.

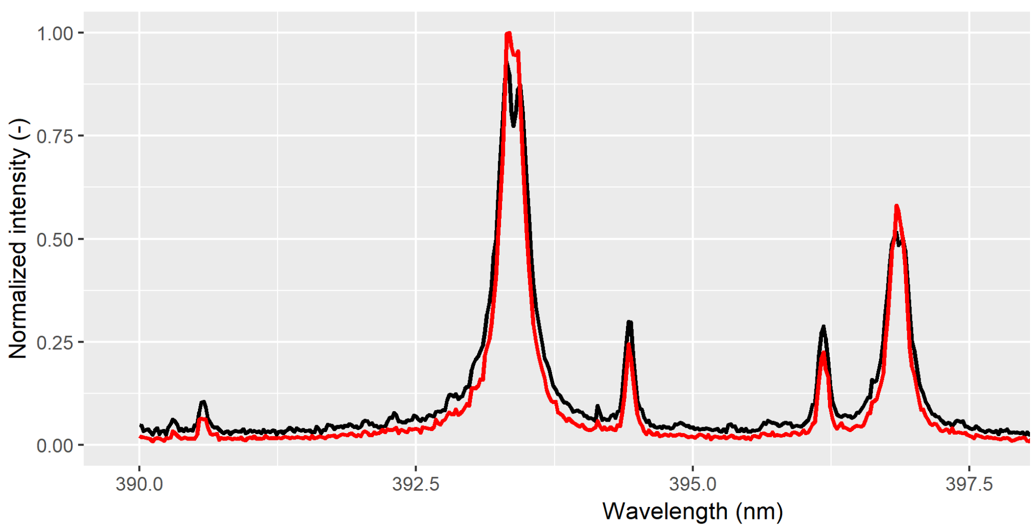
Pri RBM prakticky neexistuje hardwarový limit, keďže tréning prebieha v mini-dávkach dát. Navyše sa výpočet dá vhodne paralelizovať a rapídne zrýchliť použitím grafických kariet.

PCA algoritmus potrebuje všetky dáta naraz, a teda vzniká problém s operačnou pamäťou a manipuláciou s veľkými súbormi.

Ako limitný dátový súbor pre PCA by sa dalo považovať 10 000 echelle spektier, so 40 000 vlnovými dĺžkami.

3. otázka oponenta

Str. 52: Uvádíte, že rekonstrukci spektra užitím RBM lze považovat za úspěšnou co do poměrů a poloh čar ve spektrech. Jak byste zhodnotil změny intenzity pozadí, velikost jeho šumu a změny poměrů zájmových čar k pozadí? Z obrázků celých spekter to bez výřezů a zvětšení nelze posoudit.



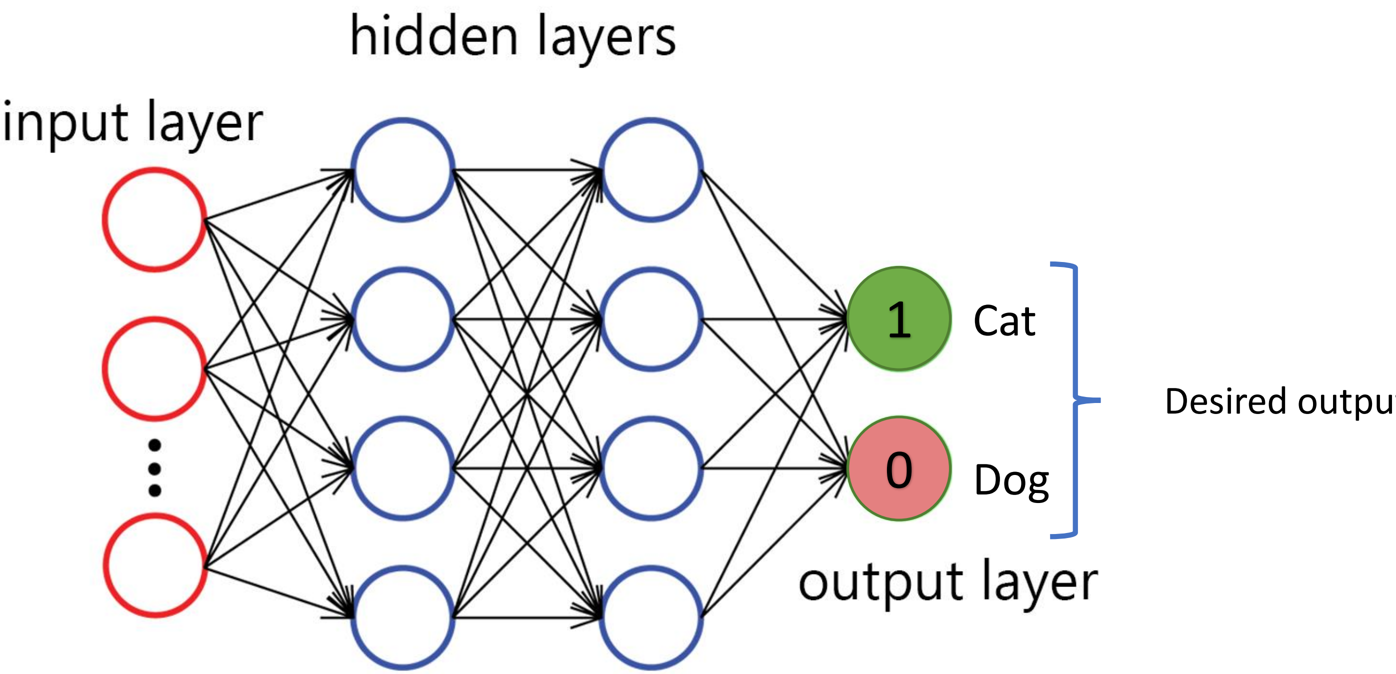
- S/B pomer mierne zhoršený
- Šum silne závisí od počtu neurónov v skrytej vrstve a od preučenosti/nepreučenosti modelu
- Pomer významných čiar zachovaný

4. otázka oponenta

Str. 57: Můžete uvést příklad opravy narušených dat generací nového spektra pomocí RBM? Může být z literatury i z Vašich experimentálních dat.

- J. Xie, 2012: Image Denoising and Inpainting with Deep Neural Networks

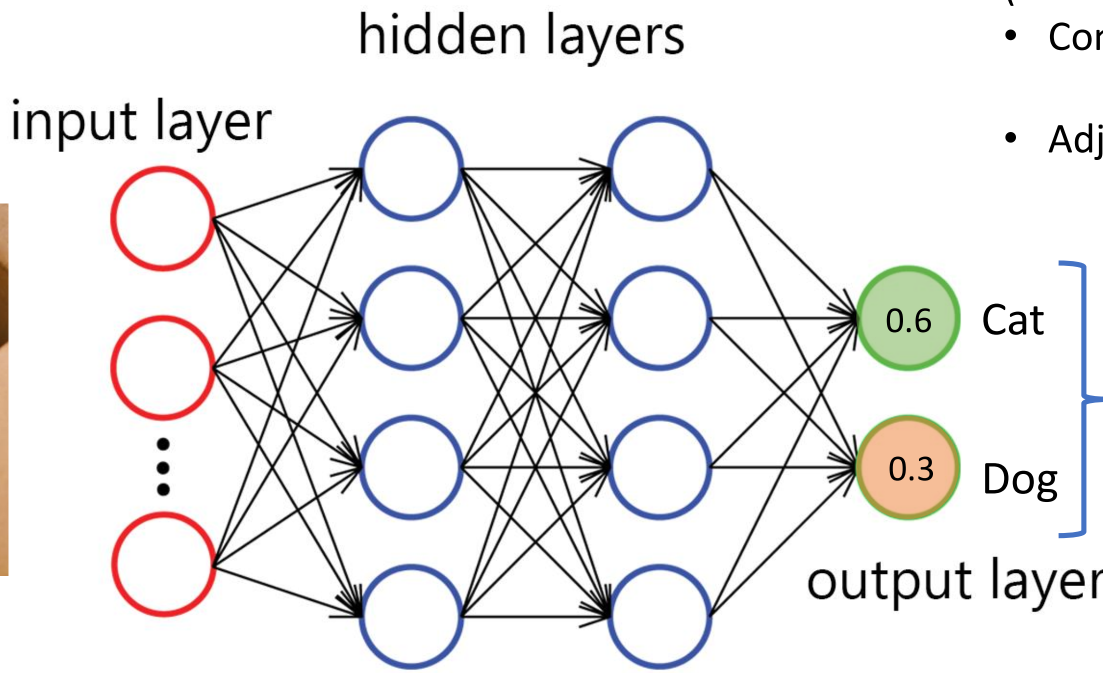
Neurónové siete – s učiteľom



$\xi = w_0 + \sum_i w_i x_i$ ...inner potential

$\sigma(\xi) = \frac{1}{1+e^{-\xi}}$ ... activation function (non-linear)

Neurónové siete – s učiteľom

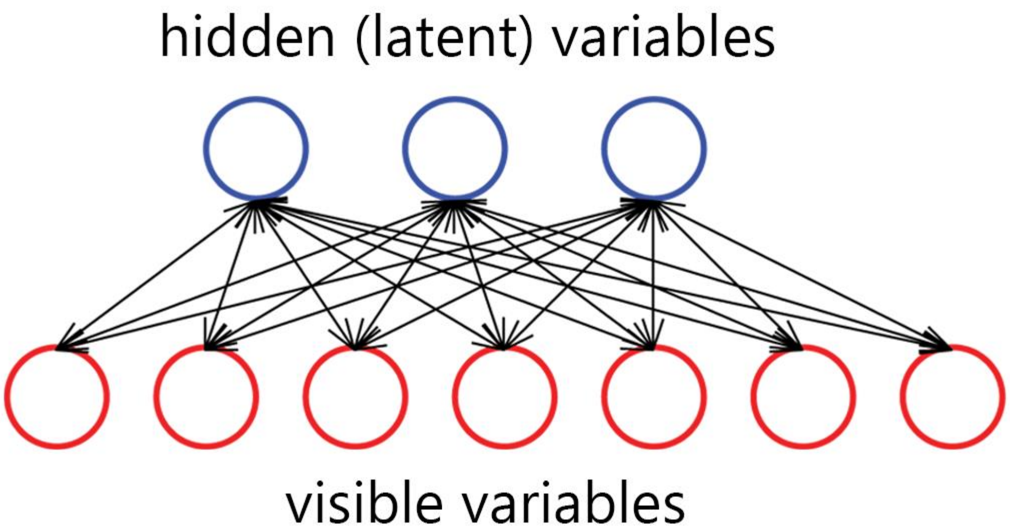


Cost function > 0
(mean squared error)

- Compute gradient: $\nabla C(w_i)$
- Adjust parameters

Real output

Restricted Boltzmann Machine (RBM)



Learning probability distr.:
- Gibbs sampling

Latent (hidden) variables
- Capturing correlations
- Dimension reduction

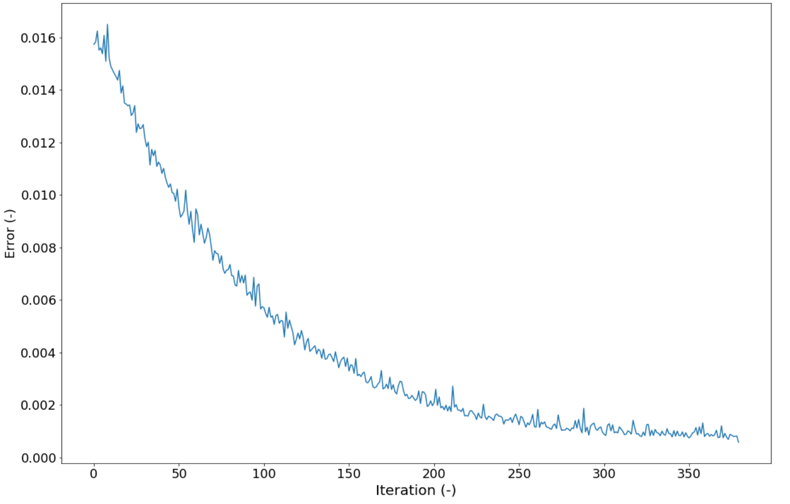
$$E(\mathbf{v},\mathbf{h}) = -\sum_i a_i v_i - \sum_\mu b_\mu h_\mu - \sum_{i\mu} w_{i\mu} v_i h_\mu$$

Approximate distribution P with a simpler distribution Q

$$\beta F_\theta = D_{KL}(Q||P) + \beta F$$

$$D_{KL}(Q||P) = \sum_x Q_\theta(x) \ln \frac{Q_\theta(x)}{P(x)}$$

- At critical points: I the correlation length of the system diverges I system becomes scale invariant I the properties of the system are characterized by critical exponents Many disparate physical systems have the same critical exponents, this is known as universality.



- Hopfield networks
  - suffer from spurious local minima that form on the energy hypersurface
  - require the input patterns to be uncorrelated
  - are limited in capacity of patterns that can be stored
  - are usually fully connected and not stacked