First meeting of Deep Lunch Study Group

Machine Learning Basics

October 19, 2016

Pham Quang Nhat Minh (Tokuda Minh)
FPT Technology Research Institute (FTRI)

Outline

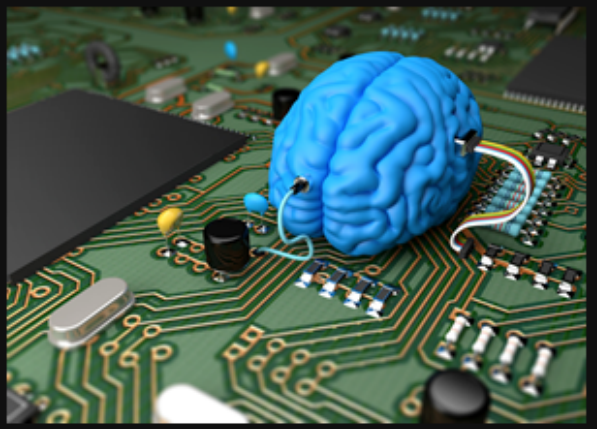
- What is Deep Lunch?
- Why Deep Lunch?
- How does Deep Lunch work?
- Machine Learning Basics
- Vocabulary for Deep Learning
- Practice exercises
  - Implement Perceptron algorithm
  - Install Theano
  - First example with Theano

What is Deep Lunch?

- A Deep Learning study group @FPT
- Study basics about Deep Learning
  - Both theory and practice aspects
- Offline meeting at lunch time
- Peer-to-peer study group
  - Teach/explain topics for others
  - Learn from others

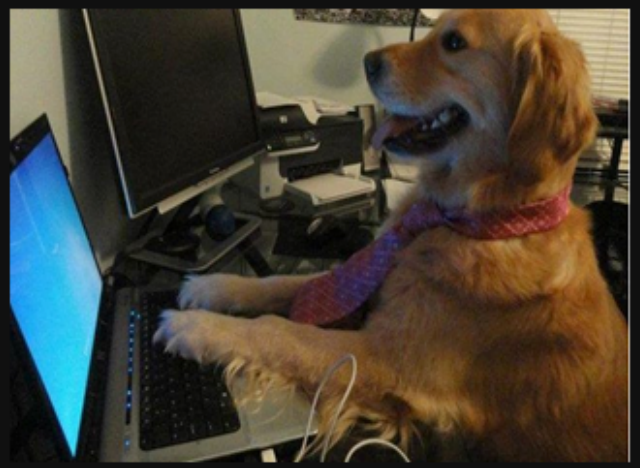
Deep Learning

What society thinks I do



What my friends think I do

What other computer scientists think I do



What mathematicians think I do

What I think I do

```
from theano import *
```

What I actually do

Why Deep Lunch?

- Deep Learning is currently hot topic in Machine Learning
- Everyone talks about Deep Learning but…
  - Not everyone really understands it
- Beyond using deep learning library
  - Know and understand basics
  - Understand how it works and how it does not work
  - Use deep learning more efficiently
  - Can modify and extend functions of deep learning library

How to NOT do machine learning

1) don't look at your data
2) don't try to interpret your results, because NNets
3) don't investigate classic approaches (physics, features? yiekes! right?)
4) start with the hardest: Deep RL and/or RNNs in vaguely defined problems
5) believe life is a data science competition w/ pre-proc data
6) forget about the implicit assumption that training and
test data follows the same distribution when that assumptiom holds FALSE
7) motivate your totally unrelated problem and assumptions with the
success of Alpha GO
8) use Terminator pictures whenever you mention "learning"
9) stop reading about how your peers are atacking same problems
10) follow up by skiping any paper without "deep" in the title
11) if all the above doesn't feel like enough post your problem
in an open source forum asking people to solve it for you

The image is credited by Eder Santana @edersantana

How does Deep Lunch work?

- Weekly offline meeting (or *once per two weeks?*)
- A member will be a chairman in each offline meeting
- Participants read materials about assigned topics and teach/explain to others
- We are neither teachers nor students
  - Or we are both teachers and students

Discussion

- Should we do weekly meeting or less frequent?
- Study level: should we study in details almost every concepts of deep learning or just understand their main ideas?
  - The former way requires longer time to study but may strengthen your knowledge more than the latter way.
- Use Theano or Tensorflow or both?
  - Theano: more easily to customize networks at low-level
  - Tensorflow: *I do not have information*

Outline

- What is Deep Lunch?
- Why Deep Lunch?
- How does Deep Lunch work?
- Machine Learning Basics
- Vocabulary for Deep Learning
- Practice exercises
  - Implement Perceptron algorithm
  - Install Theano
  - First example with Theano

Machine Learning Basics

- Definition of machine learning
- Linear Regression
- Logistic Regression
- Softmax regression
- Stochastic Gradient Descent

Definition of machine learning

- Tom Mitchell (1997): “A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by P, improves with experience *E*.”
- Three components:
  - Task *T*
  - Performance measure *P*
  - Experience *E*

The task *T*

- **Classification**
  - Specify which of *k* categories some input belongs to
  - We need to produce a function $f : R^n \rightarrow \{1, \ldots, k\}$
  - E.g., classify a news article into {*sport, education, politics,…*}, object recognition,…
- **Classification with missing inputs**
- **Regression**: predict a numerical value given some input

The task *T*

- **Transcription**: transform some kind of data into discrete, textual form
  - Optical character recognition
  - Process address numbers (Google Street Views)
  - Speech recognition
- **Machine Translation**
- **Structured output**
  - Several values that are tightly inter-related
  - Syntax Parsing, Image caption generation

The task *T*

- **Anomaly detection**
- **Synthesis and sampling**
- **Imputation of missing values**
- **Denoising**
- **Density estimation or probability mass function estimation**
  - Learn probability density function (or mass function) from data

The Performance Measure, *P*

- Depend on the task *T*
- Classification, Classification with missing inputs, Transcriptions
  - Accuracy
  - Error rate
- Density estimation
  - Average log-probability
- Use test set for evaluation
  - Test set is separated from training data

The Experience, *E*

- Experience *E* is in general in the form of a data set
  - A collection of **examples** (**data points**)
- Machine learning algorithms can be categorized into *supervised* and *unsupervised* learning based on what kind of experience we have
- Unsupervised learning algorithms
  - A data set contains many features and we learn useful properties of the data set
- Supervised learning algorithms
  - Each example is associated with a **target** or a **label**

The Experience, *E*

- The experience *E* may be not in the form of a fixed data set
  - Reinforcement learning algorithms
- Representation of a data set
  - **Design matrix:** different example in each row, each column corresponds to a different feature
  - **A set of elements:** feature vectors may have variant sizes

Example: Linear regression

Living area (feet^2)	#bedrooms	Price (1000$s)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
...	...	...

Example: linear regression

- Input: a feature vector x in R^n
- Output: y is a linear function of input
- The task, T: is to predict y from the input x
- The performance measure, P
  - Mean square error of the model on the test set

Machine Learning Basics

- Definition of machine learning
- Linear Regression
- Logistic Regression
- Softmax regression
- Stochastic Gradient Descent

Linear regression

- Part I, cs 229 lecture note: http://cs229.stanford.edu/notes/cs229-notes1.pdf
- Practice exercise: http://tinyurl.com/hckz2rr

Machine Learning Basics

- Definition of machine learning
- Linear Regression
- Logistic Regression
- Stochastic Gradient Descent
- Softmax regression

Logistic Regression

- See the lecture 6: logistic regression (Andrew Ng's machine learning course on cousera)
- Practice: Classifying MNIST digits using Logistic Regression
  - http://deeplearning.net/tutorial/logreg.html#logreg

Stochastic Gradient Descent

- First, we need to understand the basic gradient descent algorithm:
  - See the lecture of Andrew Ng, linear regression and logistic regression.
  - Other reading:
    - https://spin.atomicobject.com/2014/06/24/gradient-descent-linear-regression/
    - http://cs231n.github.io/optimization-1

Stochastic Gradient Descent

- Suggested readings:
  - Learning with large data sets (video lecture), by Andrew Ng: http://tinyurl.com/jcc4ycv
  - On overview of gradient descent optimization algorithms, by Sebastian Ruder: http://tinyurl.com/zf7aqsd
  - Section 5.9, Stochastic Gradient Descent: http://www.deeplearningbook.org/contents/ml.html
- Libraries/packages:
  - Stochastic Gradient Descent (SGD) library in scikit-learn: http://scikit-learn.org/stable/modules/sgd.html

Stochastic Gradient Descent

- Batch gradient descent

Repeat until convergence {

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} \qquad \text{(for every } j\text{)}.$$

}

- Stochastic Gradient Descent

Loop {

for i=1 to m, {

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} \qquad \text{(for every } j\text{)}.$$

}

}

Machine Learning Basics

- Definition of machine learning
- Linear Regression
- Logistic Regression
- Stochastic Gradient Descent
- Softmax regression

Outline

- What is Deep Lunch?
- Why Deep Lunch?
- How does Deep Lunch work?
- Machine Learning Basics
- Vocabulary for Deep Learning
- Practice exercises
  - Implement Perceptron algorithm
  - Install Theano
  - First example with Theano

Machine Learning Basics

- Definition of machine learning
- Linear Regression
- Logistic Regression
- Stochastic Gradient Descent
- Softmax regression

Vocabulary for Deep Learning

See slide of Graham Neubig:
https://www.phontron.com/slides/neubig14deeplunch11.pdf

- Prediction problems
- Linear classifiers
- Feature function
- Perceptron algorithm
- Sigmoid function (Logistic function)
- Logistic regression
- Stochastic gradient descent
- Gradient of the sigmoid function

Vocabulary for Deep Learning

See slide of Graham Neubig:
https://www.phontron.com/slides/neubig14deeplunch11.pdf

- Neural networks
- Multi-layer perceptrons
- Hidden Layer Activation Functions
- Deep Networks
- Back-propagation algorithm
- Vanishing Gradients
- Layerwise training
- Autoencoders

Vocabulary for Deep Learning

See slide of Graham Neubig:
https://www.phontron.com/slides/neubig14deeplunch11.pdf

- Dropout
- Recurrent Neural Networks
- Convolutional Neural Networks
- Recursive Neural Networks
- Deep belief networks
  - Restricted Boltzmann machines
  - Contrastive estimation
- Batch/mini-batch update strategies
- Long short-term memory (LSTM)
- Learning on GPUs

Outline

- What is Deep Lunch?
- Why Deep Lunch?
- How does Deep Lunch work?
- Machine Learning Basics
- Vocabulary for Deep Learning
- Practice exercises
  - Implement Perceptron algorithm
  - Install Theano
  - First example with Theano