# Predicting Disease Progression using BiDirectional LSTM


Promotor: Roel Wuyts

Milan van der Meer

# Overview

- Motivation
- Problem Definition
- Approach
  - Irregular Intervals
  - Standarization
  - High Dimensionality
  - Prediction
  - Training Model
- Dataset
- Libraries

# Motivation

# Motivation

Prerequisite precision medicine:

- Current patient state
- Disease progression
- Relation outcome disease progression to medicine

Machine Learning techniques on large-scale database:

- Modern ML techniques
- THIN database

Visualize disease progression as Graph Model

# Problem Definition

# Problem Definition

Medical history as time series:

- Medical entry = data point in time
- Long time periods
- Irregular intervals
- Standarization of attributes
- Missing values
- High dimensional
- Prediction

# Problem Definition

Time series:

- Each patient is an independent time serie
- One patient can have independent disease periods
- Ignore genetic connections between patients

Long time periods:

- Long range of dependencies
- Decay or blow up of input events (vanishing gradient)

# Problem Definition

Irregular Intervals:

- = missing data
- Tranform irregural intervals to regular intervals

Standarization of attributes

Missing values ~ High dimensionality

# Problem Definition

High Dimensional:

- Curse of Dimensionality
- Sparse data
- More data needed

Prediction:

- See previous problems
- A lot of different methods

# Approach

# Approach: Irregular Invervals

# Irregular Intervals

Irregular intervals seen as missing data:

- Interpolate
  - Non-trivial for medical data
- Predict missing values
  - Imputation
  - Also using ML-techniques
- Introduce artificial ‘empty’ value
  - Easy
  - Also a sort of information

# Approach: Standarization

# Standarization

Disease names:

- ICD format
- Already applied on database
- Possibility to generalize disease categories

Measurements:

- Standard in all areas
- Check for consistency

# Approach: High Dimensionality

# High Dimensionality

Curse of Dimensionality:

- More data is needed
- Medical data → outliers are important

LSTM:

- Papers handle HD often
- Not that much of a problem as standard algorithms

→ Still a problem though

→ But because they keep history, it is doable

# Approach: Prediction

# Sequence Labelling

Input:

- Sequence of states (diagnoses)
- Each state has a label (ICD)
- Each state has attributes

Output:

- Sequence of states including future states
- Past labels can be changed (faulty diagnose)

# Sequence Labelling

Definitions:

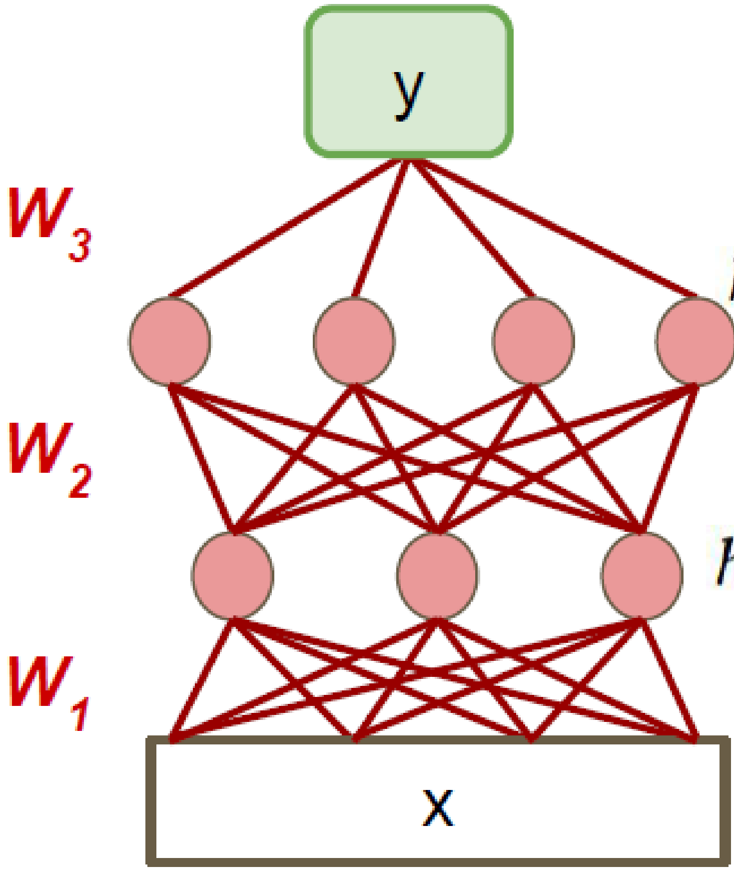
- Sequence Classifying (different patients)
- Temporal Classifying (different stages for 1 patient)

Alternative:

- Predict influence of medication/treatment
- Indirect from previous definitions

Method → Neural Network

# Neural Networks



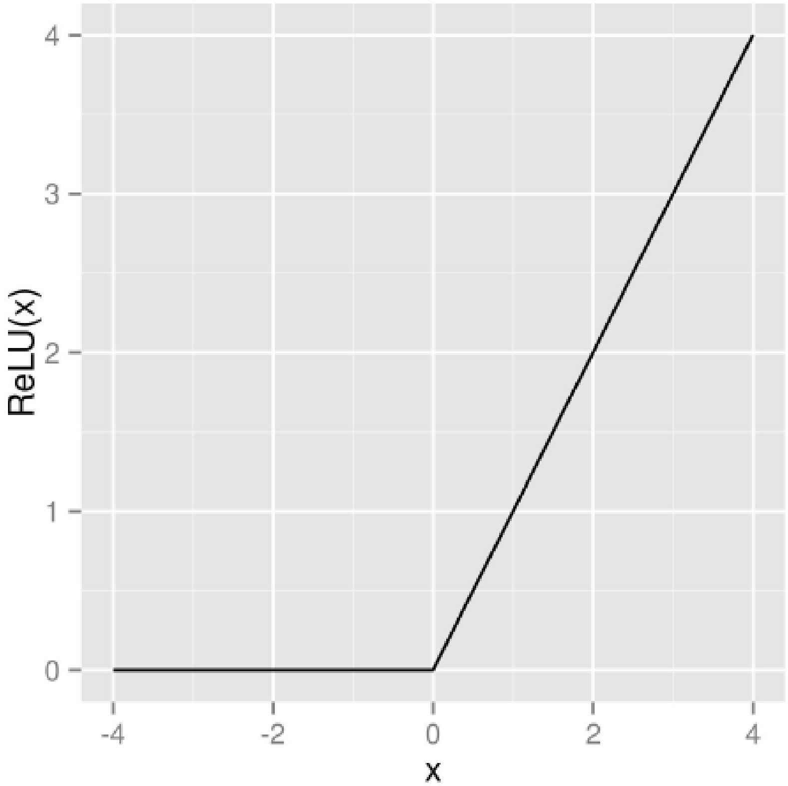

$\hat{y} = W_3 h_2$

$h_2 = \sigma(W_2 h_1)$

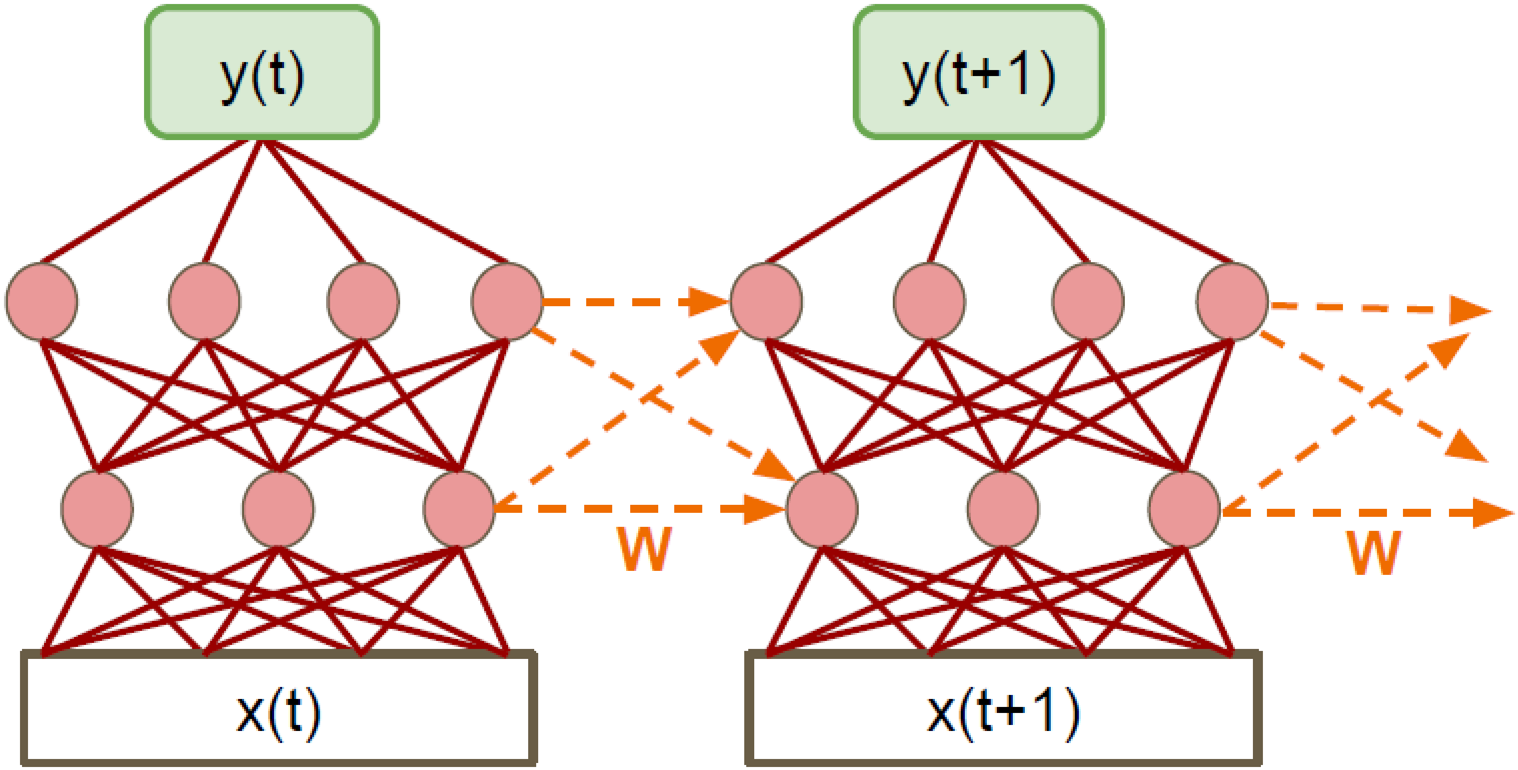
$h_1 = \sigma(W_1 x)$

# Neural Networks – Activation Function

**ReLU**:
also good accuracy
computationally cheap



# Recurrent Neural Networks



# Long Short Term Memory

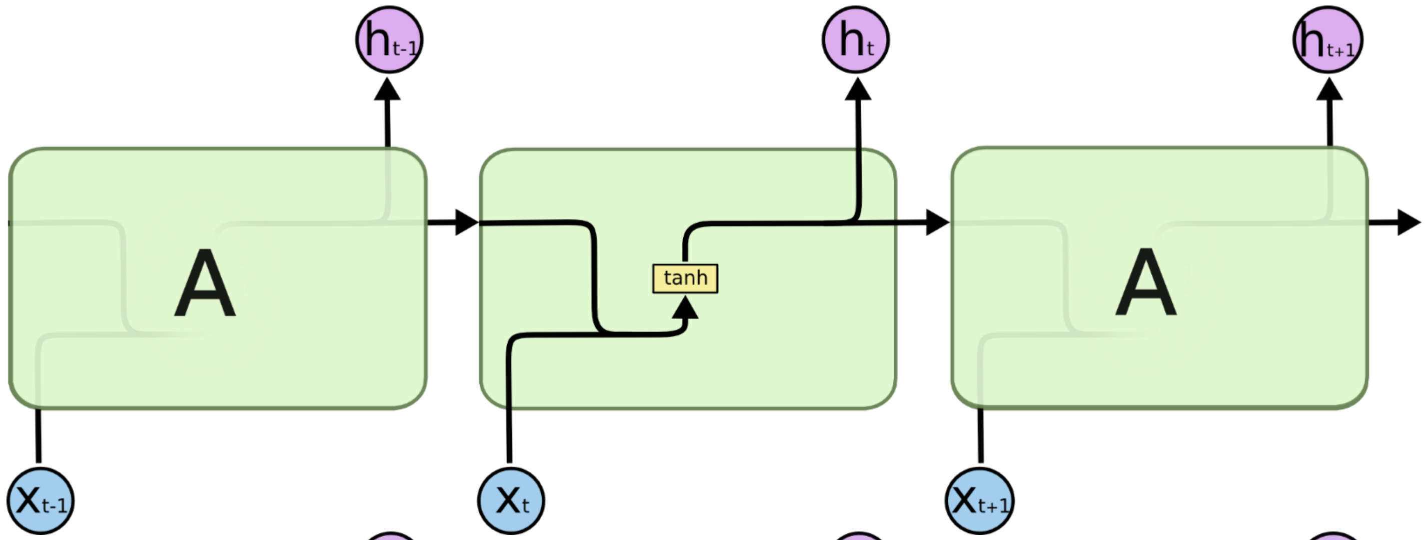
RNN:

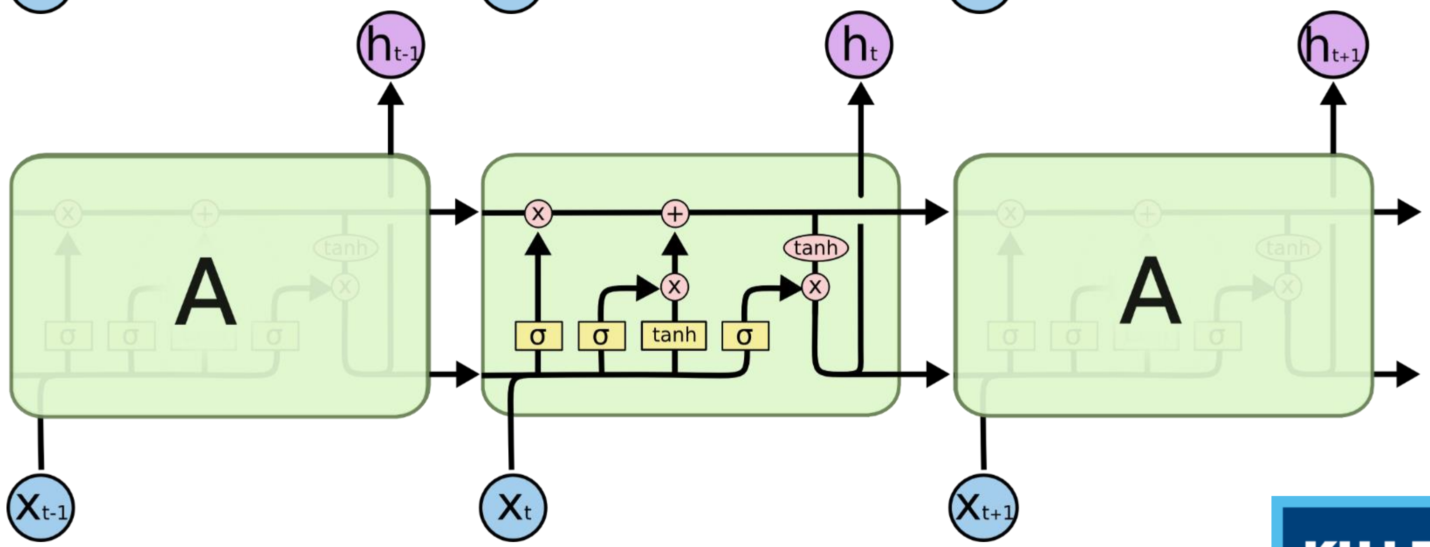
- Copies of same Neural Network
- Problem with Long-Term Dependencies
- Vanishing Gradient

LSTM:

- Designed to remember
- Work well on variety of problems

# Long Short Term Memory

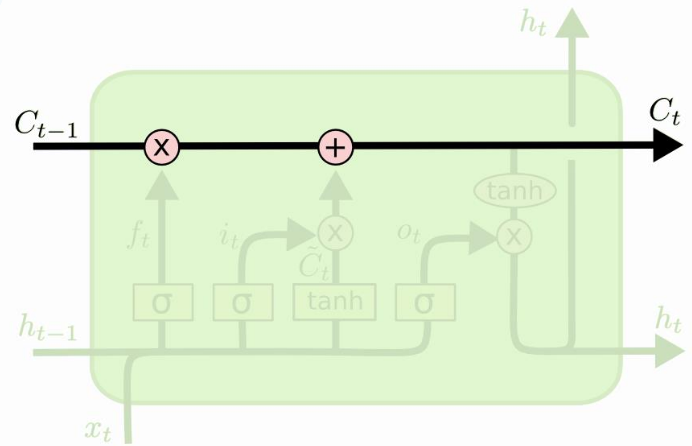


# Long Short Term Memory

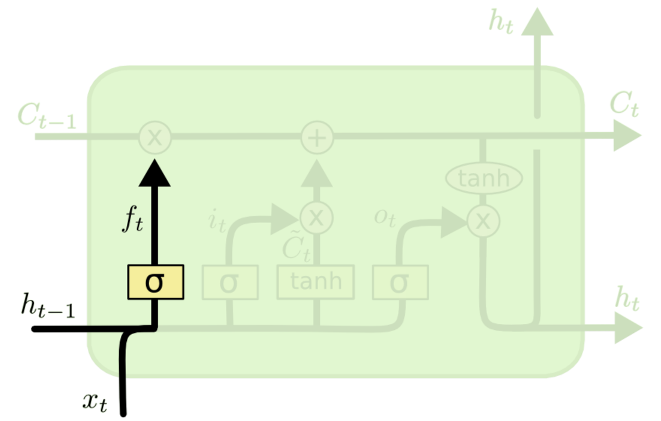
Cell state:

- Conveyor belt
- Gates




Forget gate:

- Sigmoid
- Based on input and previous layer

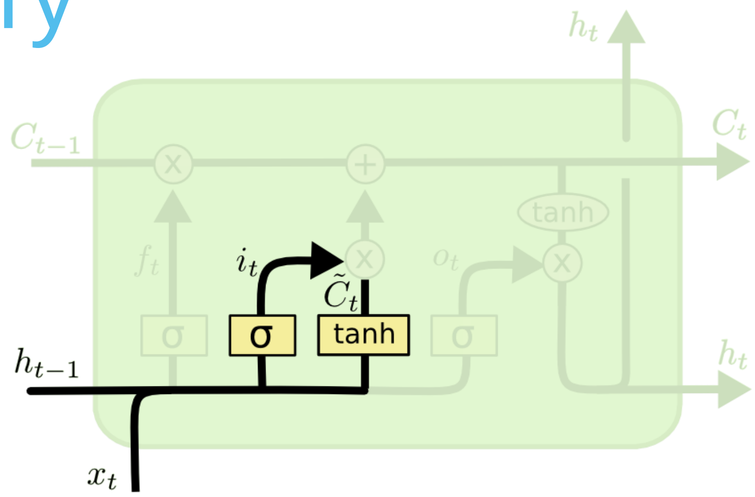



KU LEUVEN

# Long Short Term Memory

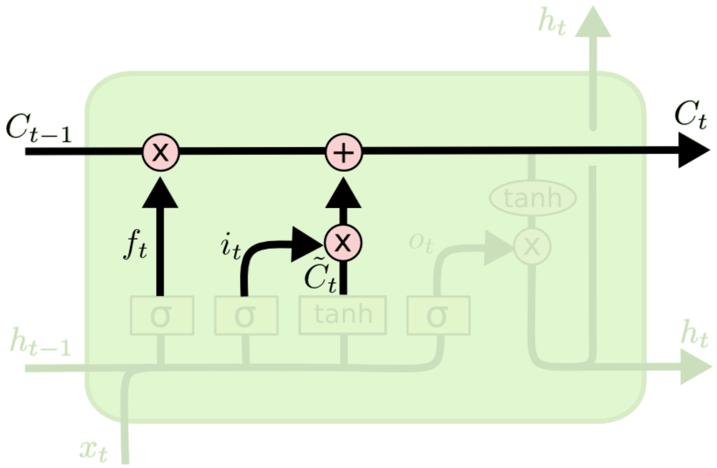
Update gate:

- Which values to update
- New candidate values




Cell state:

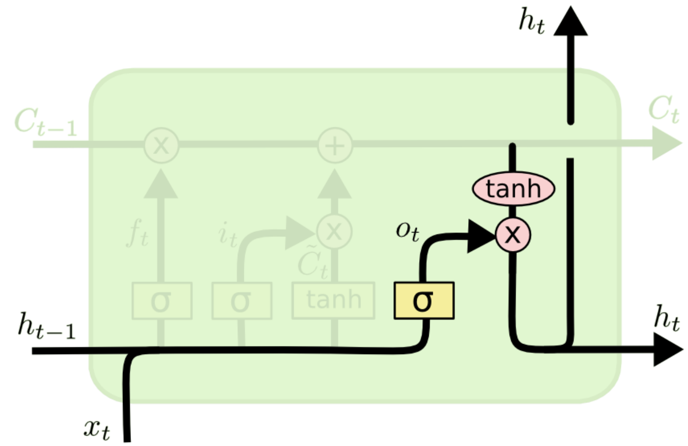
- Update



# Long Short Term Memory

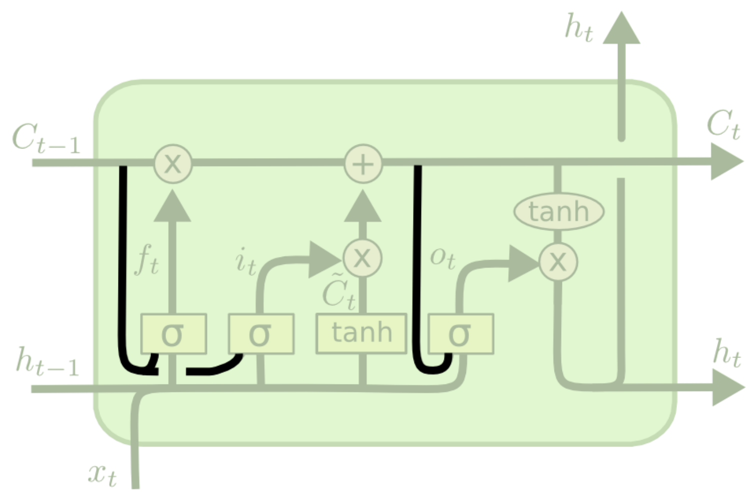
Output:

- Filter cell state
- Sigmoid
  - Decide which values




Peepholes:

- Gates can access cell state



# Long Short Term Memory

A lot of varients:

- All work about the same
- Forget gate and activation function most important
- So go for simplicity

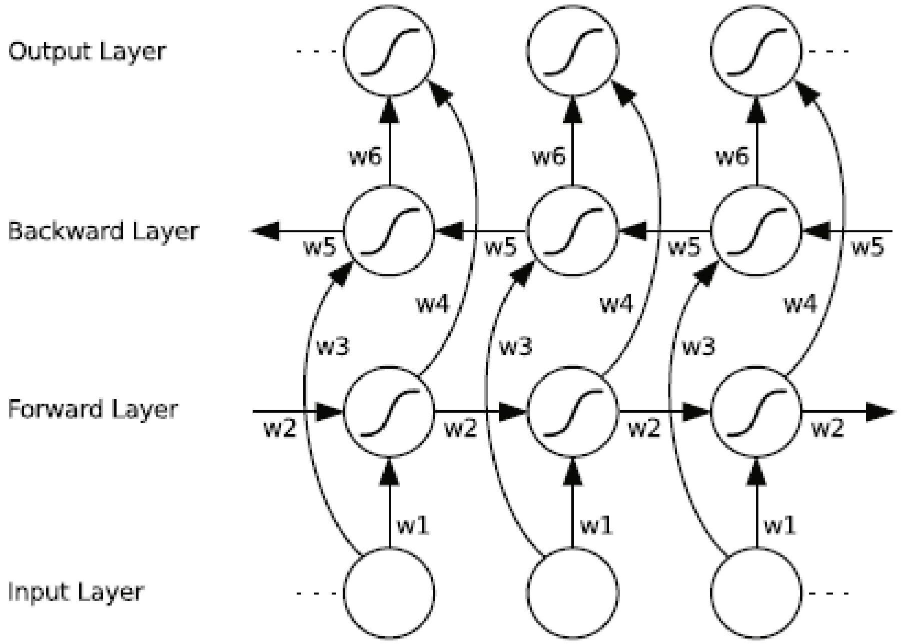
Current developments:

- Grid LSTM
- Attention LSTM

# BiDirectional Long Short Term Memory

BiDirectional:

- Access to complete sequence
- Use future context
- Applied on LSTM



# Hidden-Markov Model

HMM:

- System assumed to be a Markov Process
- But with hidden states
- Only output is vissible

Application:

- When intervals are not known
- Model sequence of states
- Convert temporal sequence to a label sequence
- Input for BLSTM

# Approach: Training Model

# Loss Function

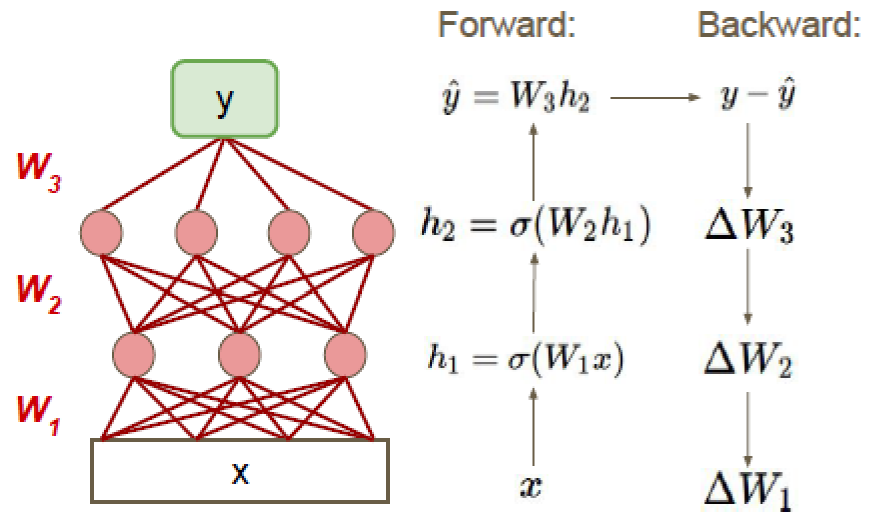
Minimizing loss function:

- Minimizing dependent on weights
- Variety of choices (standard: quadratic error)
- Depending on problem definition
  - Known intervals vs Unknown intervals
- Solved by Gradient Descent

Gradient Descent:

- Multiple methods
  - Batch
  - Online

# 2 phase algorithm




Forward Pass:

- Input x
- Calculate hidden layers
- Output y

Backward Pass:

- Calculate loss function
- Derivatives in respect to each weight
    - Require activation function to be differentiable
- Adjust each weight

# Dataset

# Dataset

THIN:

- Pseudonymised medical records
- 450 GP practices
- 3,5 million active patients
- UK population

# Libraries

# Libraries

Torch7:

- Lua
- Provides BiDirectional LSTM
- Large community

Theano:

- Python
- Provides BiDirectional LSTM
- Flexible
- Complex

# Libraries

DeepLearning4J:

- Java
- No BiDirectional LSTM

JANNLab

- Java
- BiDirectional LSTM
- Last update 2 years ago

# Libraries

RNNLIB:

- C++
- BiDirectional LSTM
- Not much documentation

# References

# References

[1] Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients

[2] Employing time-series forecasting to historical medical data: an application towards early prognosis within elderly health monitoring environments

[3] Supervised Sequence Labelling with Recurrent Neural Networks

[4] International Classification of Diseases (ICD)

[5] Phenotyping of Clinical Time Series with LSTM Recurrent Neural Networks

# References


[6] IMEC Technical talk by Jaak Simm

[7] Missing covariate data in medical research: To impute is better than to ignore

[8] Recurrent Neural Networks for Missing or Asynchronous Data

[9] Development of a Database of Health Insurance Claims: Standardization of Disease Classifications and Anonymous Record Linkage

[10] Machine Learning Based Missing Value Imputation Method for Clinical Dataset

# References


[11] Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription

[12] Long short-term memory neural network for traffic speed prediction using remote microwave sensor data

[13] Noisy Time Series Prediction using a Recurrent Neural Network and Grammatical Inference

[14] Neural Networks for Time Series Processing

[15] Constructing a Non-Linear Model with Neural Networks for Workload Characterization

# References

[16] http://colah.github.io/posts/2015-08-Understanding-LSTMs

[17] A Critical Review of Recurrent Neural Networks for Sequence Learning

[18] Bidirectional Recurrent Neural Networks

[19] Optimization of Hidden Markov Models and Neural Networks

[20] LONG SHORT-TERM MEMORY

[21] LSTM: A Search Space Odyssey

KU LEUVEN

# References

[22] Grid Long Short-Term Memory

[23] Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

[24] Hybrid HMM/ANN Systems for Speech Recognition: Overview and New Research Directions

KU LEUVEN