

# Generera rubriker med deep learning

DATX02-16-27

---

**Rickard Lantz   Jacob Genander**

Alex Evert   Nicklas Lallo   Filip Nilsson

24 februari 2016

Chalmers Tekniska Högskola

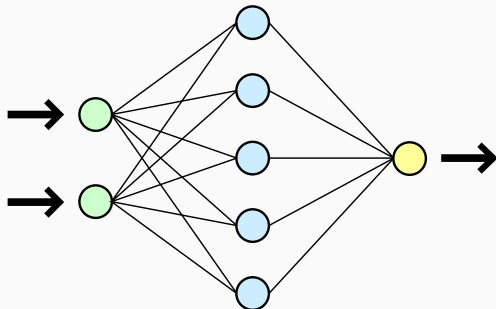
# Introduktion

---

Generering av:

1. Slumpmässiga men "vettiga" rubriker
2. Rubriker baserade på artiklar
3. Rubriker baserade på artiklar i kombination med bilder

# Artificiella Neuronnät



1

- Metod med inspiration från den mänskliga hjärnan
- Kan hitta mönster i mycket komplexa data
- Nätverket identifierar mönster som experter tidigare behövt representera i kod.

---

<sup>1</sup>[https://commons.wikimedia.org/wiki/File:Neural\\_network.svg](https://commons.wikimedia.org/wiki/File:Neural_network.svg)

## Projektets faser

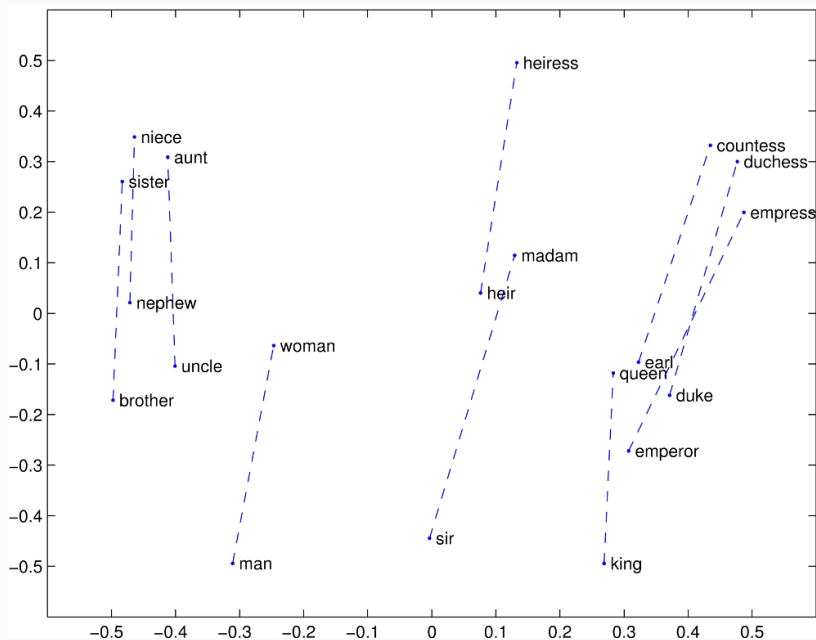
---

- Datainsamling och bearbetning
  - Artiklar med titlar/rubriker
  - "Word embeddings"
- Modellkonstruktion
- Modelljustering

Två sorters data:

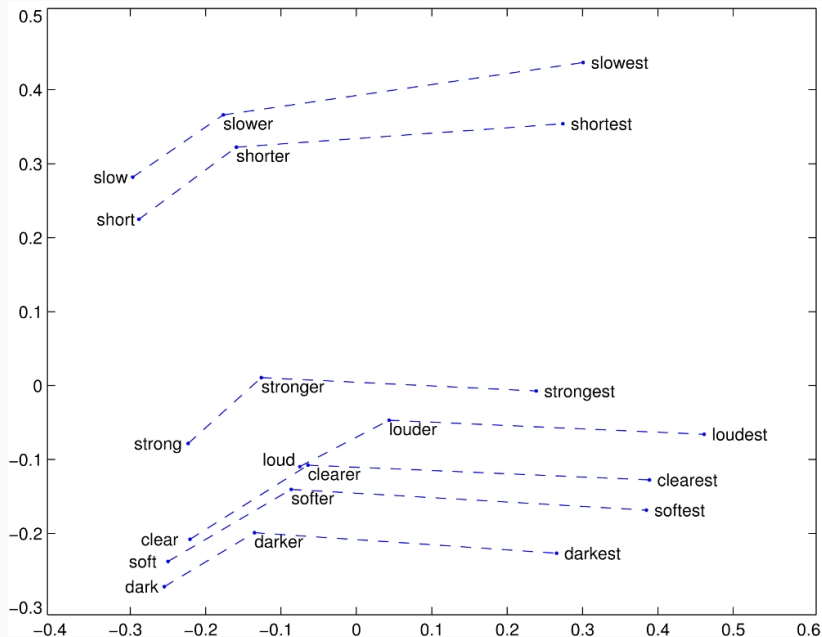
- Artiklar och titlar
  - Större dataset ger bättre generaliseringar
  - Homogent dataset
- Word embeddings
  - Neuronnät kan endast hantera numeriska data
  - Ger vektorrepresentation av ord

# GloVe – Vektorrepresentation av ord





# GloVe – Vektorrepresentation av ord



- LSTM-celler, "Long Short-Term Memory"
  - Viktar in ny data
  - Bra på sekvenser
- Ramverket TensorFlow<sup>TM</sup> från Google
  - Färdiga implementationer av ex. LSTM-celler
  - Parallell exekvering på CPU:er och GPU:er

Finns inget "facit", endast riktlinjer

- Iterativ, testbaserad process
  1. Träning av nätverket
  2. Utvärdering av nätverkets prestanda
  3. Justerande av hyperparametrar
    - Nätverkets storlek
    - Inlärningstakt
    - ...

# Sammanfattning

---

Projektets fas:

- ✓ Datainsamling och bearbetning
- ✓ Modellkonstruktion
- Modelljustering

Uppnådda mål:

- ✓ Slumpmässiga men "vettiga" rubriker
- Rubriker baserade på artiklar
- Rubriker baserade på artiklar i kombination med bilder

- Tar tid att sätta sig in i ett komplext ramverk inom ett nytt fält
- Ingen exakt vetenskap
- Verkar som magi på ytan, men går ändå att förstå på ett intuitivt plan