

F11 - Maskininlärning och AI

Statistisk översiktscurs

Ulf Högnäs

Statistiska institutionen
Stockholms universitet

April 23, 2025



Stockholm
University

Contents

- 1 Introduktion
- 2 Maskininlärning
- 3 AI och Generativ AI
- 4 Natural Language Processing och Chatbottar
- 5 Bild- och Videomodeller

Introduktion

- I den här föreläsningen får vi igenom
 - Supervised machine learning
 - Large Language Models
 - Neurala nätverk
 - AI
 - Generativ AI
- Alla dessa modeller är statistiska modeller
 - Modellerna har parametrar (jämför med β_0, β_1, \dots i regression)
 - Parametrarna anpassas efter data
 - Målsättningen är att minimera felen

Exempel på tillämpningar av ML och AI

- Bildigenkänning (t.ex. ansiktsigenkänning, medicinska bilder)
- Språkteknologi (t.ex. översättning, taligenkänning, chattbottar)
- Rekommendationssystem (t.ex. Netflix, Spotify, Amazon)
- Självkörande fordon
- Bedrägeridetektion (t.ex. inom bank och försäkring)
- Prediktivt underhåll (t.ex. i industriella maskiner)
- Diagnostik och beslutstöd inom vården
- Finansiell prognostisering (t.ex. aktiemarknadsanalys)
- Personalisering av reklam och marknadsföring
- Spelutveckling och AI-motståndare

Maskininlärning

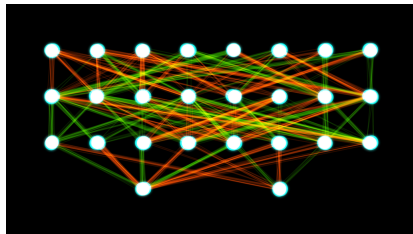
Training Data och Testing Data

Exempel: maskinell inläsning av handskrivna siffror

- Vi har 60,000 handskrivna siffror med “facit”, representerade som relativt lågupplösta pixlar
- Vi sätter 10,000 av dessa åt sidan (**testing data**, “validation data” eller “hold-out set”)
- Modellen tränas på 50,000 siffror (**training data**). Modellen ...
 - 1 gissar svaret
 - 2 mäter hur stor andel som blev fel
 - 3 modellens interna parametrar justeras
 - 4 repeat!
- När träningen är klar provas modellen på “testing set”. Om träffsäkerheten är tillräckligt bra så är vi klara.

Neurala Nätverk

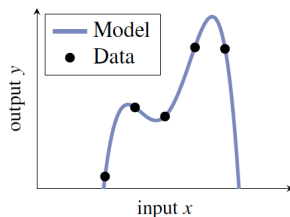
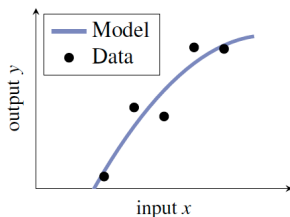
- Grunden lades på 1940-talet
- **Backpropagation** 1980-talet
- De stora tekniska genombrotten 2010-talet
- Nobelpriset i fysik, 2024
- Minikurs på Youtube, 3Blue1Brown



Av Akritasa CC BY-SA 4.0

Overfitting

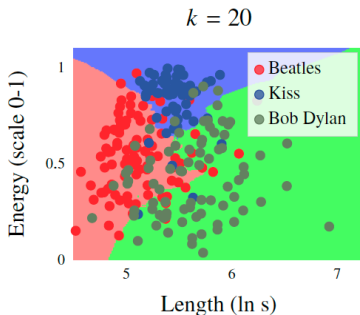
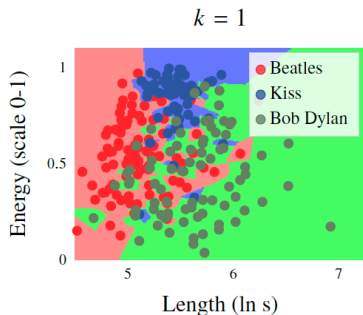
- Varför döljer man en del data i ett “testing set”?
- Modellerna i ML är väldigt flexibla
- Stor flexibilitet kan leda till **overfitting**



Vilken modell ger bättre prediktioner? Bilder från Lindholm et al. (2022)

Overfitting

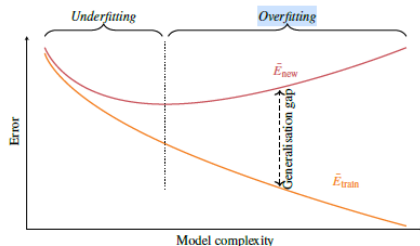
Om vi får in data för en låt där gruppen eller artisten är dold, vilken av dessa modeller har bäst chans att klassificera rätt?



k-nearest-neighbor Bilden till vänster är ett exempel på overfitting

Hyperparametrar och modellval

- 1 Träna en modell på training set
- 2 Mät hur bra modellens skattningar passar testing set
- 3 Upprepa steg 1 och 2 med andra modeller
- 4 Välj den modell som förutsäger testing set bäst



$k = 50$ leder till underfitting, $k = 1$ leder till overfitting

AI och Generativ AI

Vad är AI och generativ AI?

Detta är skrivet av Chat-GPT 4o

Artificiell Intelligens (AI)

AI är ett samlingsnamn för tekniker som gör det möjligt för datorer att utföra uppgifter som normalt kräver mänsklig intelligens, såsom att känna igen mönster, fatta beslut, förstå språk och lära sig från data.

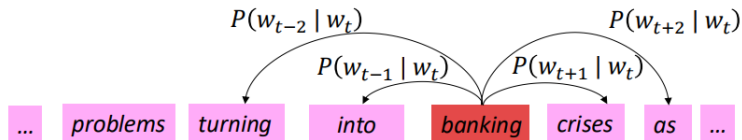
Generativ AI

Generativ AI är en gren av AI som kan skapa nytt innehåll – till exempel text, bilder, ljud eller kod – utifrån träningsdata. Exempel på generativ AI är ChatGPT, DALL·E och musikgenereringsverktyg.

Natural Language Processing och Chatbottar

Natural Language Processing och Chatbottar

- Datorer förstår siffror, inte text
- Lösningen är att representera varje ord som vektorer (ordnade nummer)
- Efficient Estimation of Word Representations in Vector Space [Mikolov et al. \(2013\)](#) “**Word2Vec**”



Skip-gram: ett sätt att förutsäga kringliggande ord med en statistisk modell

Vector Representations

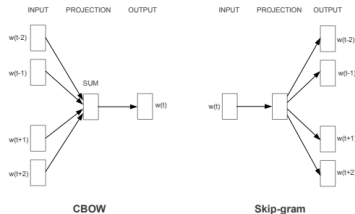


Figure 1: New model architectures. The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word.

```
[ 0.23088  0.28283  0.6318 -0.59411 -0.58599  0.63255
 0.24402 -0.14108  0.060815 -0.7898 -0.29102  0.14287
 0.72274  0.20428  0.1407  0.98757  0.52533  0.097456
 0.8822  0.51221  0.40204  0.21169 -0.013109 -0.71616
 0.55387  1.1452 -0.88044 -0.50216 -0.22814  0.023885
 0.1072  0.083739  0.55015  0.58479  0.75816  0.45706
-0.28001  0.25225  0.68965 -0.60972  0.19578  0.044209
-0.31136 -0.68826 -0.22721  0.46185 -0.77162  0.10208
 0.55636  0.067417 -0.57207  0.23735  0.4717  0.82765
-0.29263 -1.3422 -0.099277  0.28139  0.41604  0.10583
 0.62203  0.09496 -0.23446  0.51349  0.99379  1.1846
-0.16364  0.20653  0.73854  0.24059 -0.96473  0.13481
-0.0072484  0.33016 -0.12365  0.27191 -0.40951  0.021909
-0.6069  0.40755  0.19566 -0.41002  0.18636 -0.032652
-0.78571 -0.13847  0.044007 -0.084423  0.04911  0.24104
 0.45273 -0.18682  0.46182  0.089068 -0.18185 -0.01523
-0.7368 -0.14532  0.15104 -0.71493 ]
(gensimEnv) uffe@laptop:~/Tresors/python/huggingface$
```

Bild från Mikolov et al. (2013)

En vektor som representerar "cat" (GloVe)

Ord i hundradimensionellt i rum

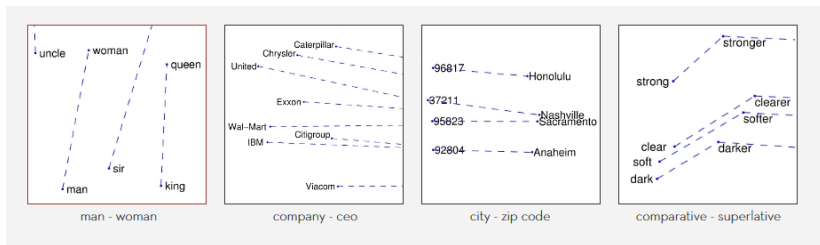
Word2vec maximizes objective function by putting similar words nearby in space



från Stanfords CS224N


King - Man + Woman \approx Queen


- Vi kan förflytta oss genom att lägga till och ta bort ord från varandra
- Sedan kan vi titta på de ord som ligger närmast den nya “platsen”
- “King - Man + Woman \approx Queen” \leftarrow berömt exempel
- Men det är sällan så intuitivt i verkligheten




från GloVe

Chat-GPT-2 (2019)

 **Inference Providers** NEW

 **HF Inference API**

 Text Generation


Examples


▼

Pet rats make good pets because they are able to survive in the wild. They are also very intelligent and can learn quickly. They

Compute

⌘+Enter

[</> View Code](#)  3.2s

 Maximize

Chat-GPT-2

Hallucinationer - 4o vs o1-preview

Integer Partitions of 11 into Three Parts:

1. $9 + 1 + 1$
2. $8 + 2 + 1$
3. $7 + 3 + 1$
4. $6 + 4 + 1$
5. $6 + 3 + 2$
6. $5 + 5 + 1$
7. $5 + 4 + 2$
8. $4 + 4 + 3$

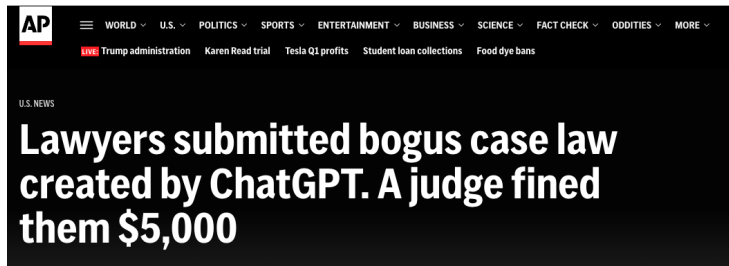
Ovan: 4o:s försök

Höger: o1-previews lösning

Integer Partitions of 11 into Three Parts:

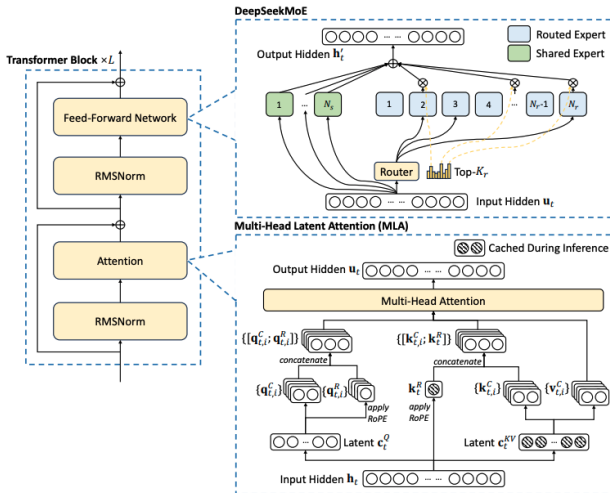
1. $9 + 1 + 1$
2. $8 + 2 + 1$
3. $7 + 3 + 1$
4. ~~$7 + 2 + 2$~~
5. $6 + 4 + 1$
6. $6 + 3 + 2$
7. $5 + 5 + 1$
8. $5 + 4 + 2$
9. ~~$5 + 3 + 3$~~
10. $4 + 4 + 3$

“Several of those cases weren’t real, misidentified judges or involved airlines that didn’t exist.”



AP June 23, 2023

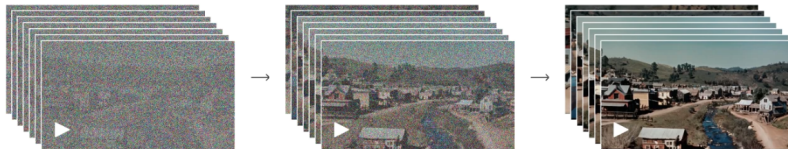
DeepSeek R1 (2024)



Deepseek. Kräver log-in (gratis)

Bild- och Videomodeller

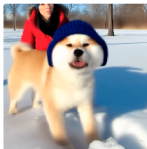
Bild- och Videoskapande



In this work, we find that diffusion transformers scale effectively as video models as well. Below, we show a comparison of video samples with fixed seeds and inputs as training progresses. Sample quality improves markedly as training compute increases.



Base compute



4x compute



32x compute

OpenAIs Sora