UiT Norges arktiske universitet

Whisper-modellen

SOK-3023 (ML for økonomer), 5 ECTS

Markus J. Aase

markus.j.aase@uit.no, kontor 02.411

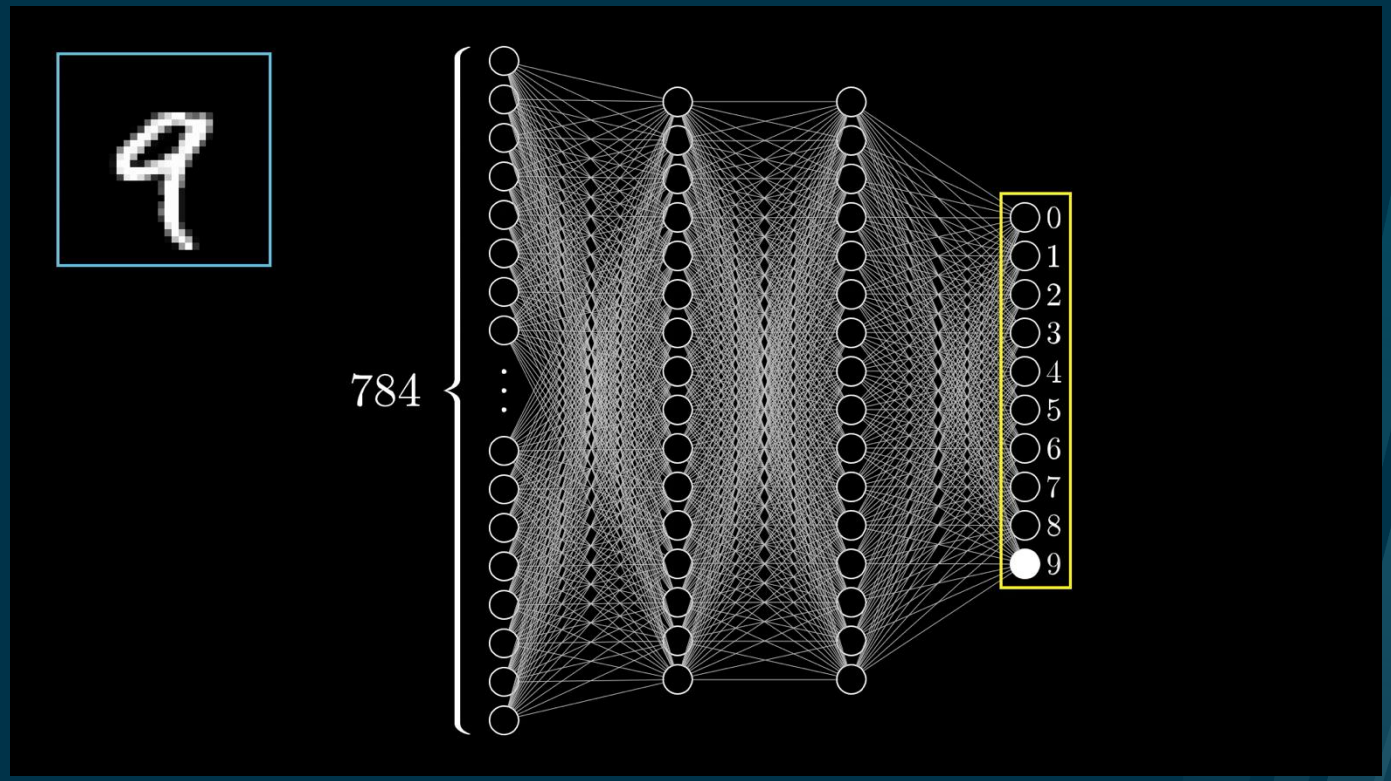
Universitetslektor i matematikk og statistikk

Handelshøgskolen, UiT

Master i samfunnsøkonomi med datavitenskap

Til nå

- Maskinlæring
  - Lineær algebra, statistikk og kalkulus
  - Prediksjon vs inferens
  - Regresjon vs klassifikasjon
  - Bias variance trade-off
  - Nødvendighet av god data
  - Reducible og irreducible error
  - Tensorer
  - Veiledet læring
  - Overfitting/underfitting
  - Trenings-, validerings- og testsett
  - Evalueringsmetrikker – accuracy, sensitivitet, precision osv.
  - Arkitektur i nevrale nettverk
  - Aktiveringsfunksjoner, loss/kost-funksjon
  - «Læring» = minimering av loss-funksjon
  - Epoch, batch, batch_size
  - Gradient descent



Videre i kurset

- Kurset er delt i 6 ulike deler
  - Intro til ML #1 – ferdig
  - Intro til ML #2 – ferdig
  - Intro til ML #3 – ferdig
  - Whisper-modellen – i dag☺
  - CNN’s – neste uke☺
  - LSTM – nesteneste uke☺

Multitask training data (680k hours)

English transcription

"Ask not what your country can do for ⋯"

Ask not what your country can do for ⋯

Any-to-English speech translation

"El rápido zorro marrón salta sobre ⋯"

The quick brown fox jumps over ⋯

Non-English transcription

"언덕 위에 올라 내려다보면 너무나 넓고 넓은 ⋯"

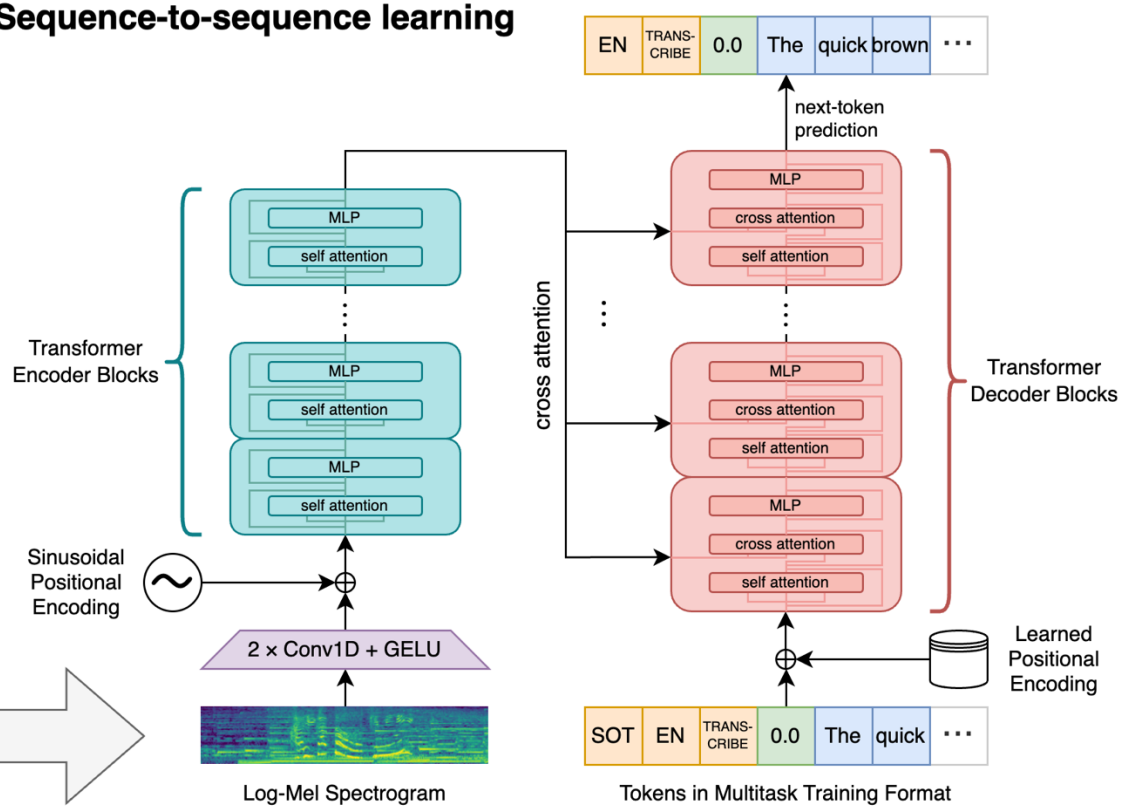
언덕 위에 올라 내려다보면 너무나 넓고 넓은 ⋯

No speech

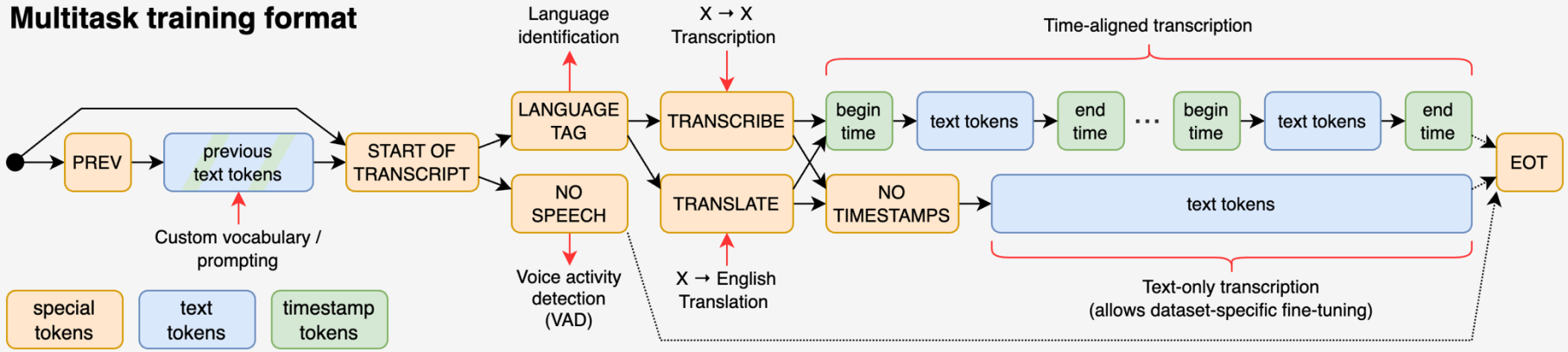
(background music playing)

∅

Sequence-to-sequence learning



Multitask training format



Hva er Whisper?

- Whisper er en automatisk talegjenkjenningsmodell (ASR – Automatic Speech Recognition) utviklet av OpenAI.
- Den er trent på store mengder flerspråklige lyddata og kan transkribere tale til tekst med høy nøyaktighet.
- Støtter over 90 språk, inkludert norsk.

Hovedegenskaper

- **Flerspråklig støtte** – Kan transkribere og forstå mange språk.
- **Robust mot støy** – Takler bakgrunnsstøy og variert lydkvalitet.
- **Tidsstempling** – Gir tidskoder til tekst, nyttig for undertekster.
- **Oversettelse** – Kan oversette tale fra et språk til engelsk.
- **Fungerer både lokalt og i skyen** – Kan kjøre på egen maskin eller via API.

Bruksområder

- Automatisert transkripsjon av møter, intervju, podkaster og forelesninger.
- Teksting av videoer og innhold på sosiale medier.
- Hjelpemiddel for hørselshemmede.
- Analyse av taleopptak i forskning og journalistikk.

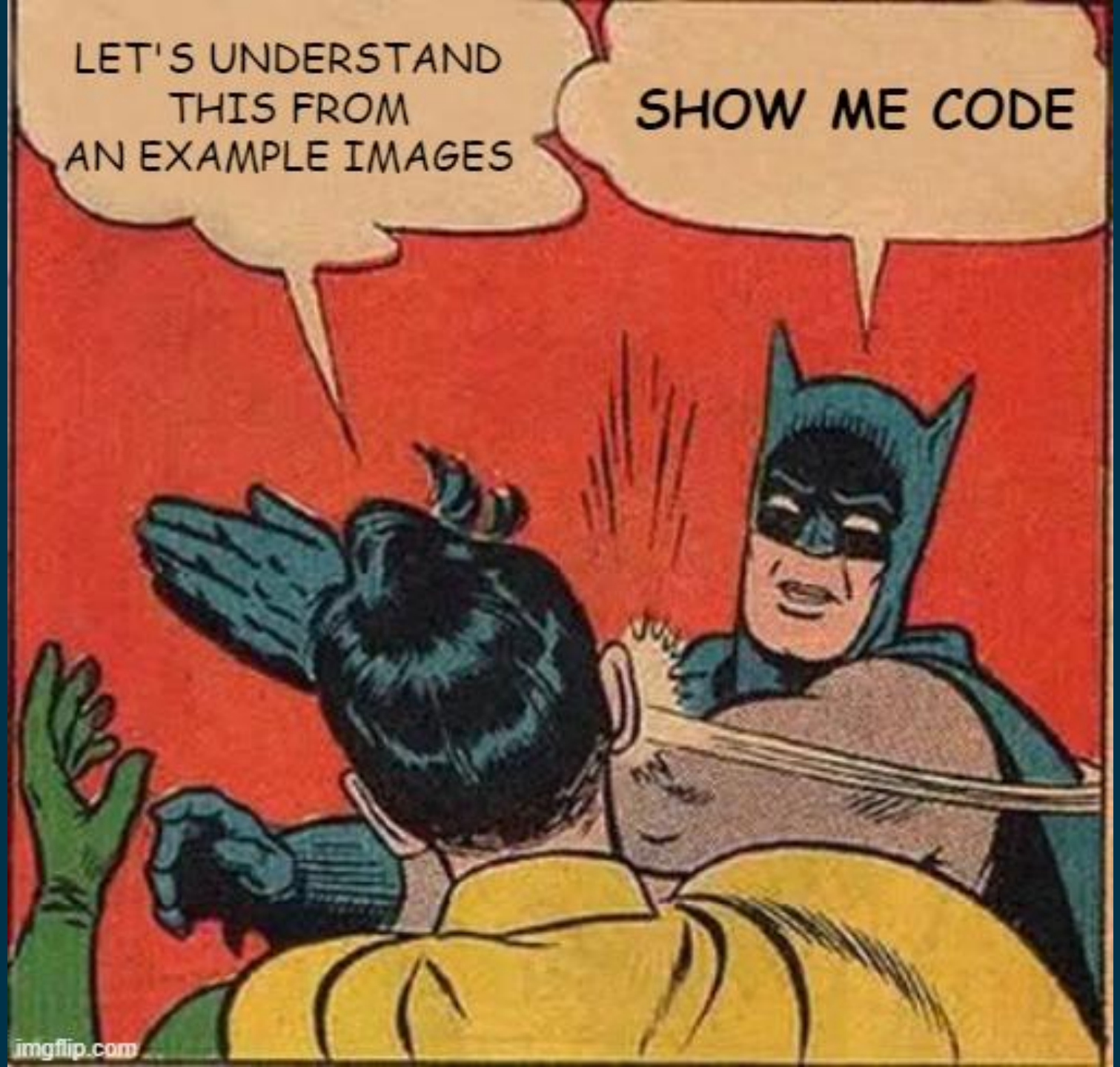
Styrker og Begrensninger

✅ Høy nøyaktighet, selv med aksenter og dialekter.
✅ Gratis og åpen kildekode.
✅ Takler lange lydklipp uten behov for oppdeling.

❌ Krever kraftig maskinvare for lokal kjøring.
❌ Kan feiltolke sjargong eller spesifikke faguttrykk.

Hvorfor er Whisper viktig?

- Setter en ny standard for talegjenkjenning med åpen tilgang.
- Gir kraftige verktøy for innholdsskaping og språkanalyse.
- Demokratiserer tilgang til avansert ASR-teknologi.


LET'S UNDERSTAND
THIS FROM
AN EXAMPLE IMAGES
SHOW ME CODE
imgflip.com