

Introduksjon

Oppgave	Tittel	Maks poeng	Oppgavetype
i	Forside		Informasjon eller ressurser

Vektorrom

Oppgave	Tittel	Maks poeng	Oppgavetype
1.1	K-means vs Rocchio	8	Langsvar
1.2	Distribusjonell semantikk	5	Langsvar
1.3	Vektor-typer	5	Langsvar
1.4	BoW + ordfrekvens	8	Langsvar
1.5	Pre-prosessering	5	Langsvar
1.6	P og R	7	Langsvar

Logistisk regresjon og HMMs

Oppgave	Tittel	Maks poeng	Oppgavetype
2.1	Logistisk regresjon	8	Langsvar
2.2	Hidden Markov Models	12	Langsvar

Dependenssyntaks og parsing

Oppgave	Tittel	Maks poeng	Oppgavetype
3.1	Dependenssyntaks	10	Langsvar
3.2	Dependensparsing	10	Langsvar

Maskinoversettelse og dialogsystemer

Oppgave	Tittel	Maks poeng	Oppgavetype
4.1	Maskinoversettelse	12	Langsvar
4.2	Chatbots	10	Langsvar

i Forside

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Skriftlig eksamen i IN2110

2022, vår

Varighet: 1. juni, 15:00 til 19:00 (4 timer)

Tillatte hjelpemidler: Ingen

Det er viktig at du leser denne forsiden nøye før du starter. Vi anbefaler også at du ser igjennom hele oppgavesettet før du starter.

1.1 K-means vs Rocchio

To av metodene vi har snakket om i forbindelse med vektorrom er *K-means* og *Rocchio*. Diskuter kort (ikke mer enn ca en halv side) disse to metodene, med vekt på forskjellene og likhetene mellom dem. Prøv å ta i bruk relevant fagterminologi.

Skriv ditt svar her

Maks poeng: 8

1.2 Distribusjonell semantikk

Forklart kort (2–3 setninger bør være nok) hva vi mener med den såkalte *distribusjonelle hypotesen* og hvordan vi kan nyttiggjøre oss av denne for å automatisk modellere betydningen til ord. Forklar også med et par setninger hva vi mener med antonymer og på hvilken måte disse kan de være en utfordring for denne hypotesen?

Skriv ditt svar her

Maks poeng: 5

1.3 Vektor-typer

I forbindelse med modellering av betydningen til ord har vi i forelesningene nevnt både tradisjonelle vektor-representasjoner og såkalte *embeddings* – hvordan skiller disse seg fra hverandre?

Skriv ditt svar her

Maks poeng: 5

1.4 BoW + ordfrekvens

Dette er en refleksjonsoppgave. Tenkt deg at vi har et system for å klassifisere tekst (setninger eller hele dokumenter) basert på en såkalt *bag-of-words* (BoW) trekk-representasjon. Vi antar videre at verdiene er rå frekvenser (opptellinger) av ord. Hvilke mulige styrker og svakheter har en slik representasjon? Diskusjonen din kan gjerne knyttes til konkrete eksempler på anvendelser. Ikke bruk mer enn ca en halv side.

Skriv ditt svar her

Maks poeng: 8

1.5 Pre-prosessering

Da vi beskrev BoW-representasjonen i foregående oppgave sa vi ingenting om nøyaktig hvordan vi definerer hva som teller som et ord. Dette er viktig for å definere trekkene i modellen, og her har vi flere design-valg tilgjengelig. Gi eksempler på hva slags pre-prosessering vi kan gjøre på teksten, og hvilke konsekvenser dette vil ha for hvor mange trekk vi får i modellen. Ikke bruk mer enn ca en halv side.

Skriv ditt svar her

Maks poeng: 5

1.6 P og R

To viktige mål for å evaluere klassifikatorer er *Precision* (P) og *Recall* (R). Disse er gjerne definert på basis av fire mer grunnleggende kategorier av hvordan prediksjonene til en klassifikator kan være riktige eller gale, sammenliknet med gullstandarden; *false positives* (FP), *true negatives* (TN), osv.

- Vis formlene for hvordan P og R er definert på basis av disse kategoriene, og forklar også med dine egne ord hva som måles (én setning på hver).
- I hvilken grad samsvarer disse målene med hverandre? Vil f.eks en høy verdi for ett av disse målene alltid bety en høy/lav verdi for det andre målet? Forklar med noen få setninger.

Skriv ditt svar her

Maks poeng: 7

2.1 Logistisk regresjon

Du har blitt bedt om å utvikle et system for å automatisk kategorisere eposter i 5 ulike klasser. Du har fått tilgang til et datasett hvor hver epost er markert med én av disse 5 klassene, og ønsker å bruke datasettet til å lære en (multinomial) logistisk regresjonsmodell.

Modellen skal ta i bruk bag-of-words som trekk, med en vokabularstørrelse på 30 000 ord.

1) Hvor mange parametre (vekker og skjæringspunkter) vil en slik logistisk regresjonsmodell inneholde?

2) I vokabularet på 30 000 ord vil mange ord forekomme sjelden, often kun én gang i hele datasettet. Kan det skape problemer for modellen? Forklar.

3) Kan regularisering brukes i et slikt tilfelle? Forklar hva regularisering handler om, og hvordan den vil påvirke parametrene i den logistiske regresjonsmodellen.

4) En av de 5 klassene handler om viktige eposter som er særdeles viktige å få riktig kategorisert. Med andre ord vil en feilklassifisering for eposter fra denne klassen ha en mye høyere kostnad enn for de andre klassene. Hvordan vil du ha endret treningen til din maskinlæringsmodell for å ta dette i betraktning?

Skriv ditt svar her

Maks poeng: 8

2.2 Hidden Markov Models

Du har utviklet en liten snakkende robot som har 3 forhåndsprogrammerte adferder:

- I den "humoristiske" adferden forteller roboten ulike vitser
- I den "alvorlige" adferden snakker roboten om viktigheten av å lese pensum
- I den "snakkesalige" adferden snakker roboten om seg selv.

Disse tre adferdene byttes ved starten av hver samtale ved å følge disse reglene:

- ved den første samtalen vil roboten ha en 40% sannsynlighet for å velge en humoristisk adferd, en 10% sannsynlighet for å velge en alvorlig adferd, og en 50% sannsynlighet for å velge en snakkesalig adferd.
- den påfølgende samtalen vil det da byttes til en av de to andre adferdene, begge to med lik sannsynlighet. Hvis roboten følger f.eks. en humoristisk adferd vil neste samtale bruke enten en alvorlig adferd eller en snakkesalig adferd, begge to med 0.5 sannsynlighet.

Du gjennomfører et eksperiment hvor deltakere går inn i et lukket rom hvor roboten befinner seg. Deltakere snakker med roboten en liten stund og forlater deretter rommet. Du kan ikke se selve roboten, men du kan derimot observere om deltakeren som går ut av rommet er enten glad eller bekymret.

Du kan anta at:

- etter en samtale med roboten som kjører en humoristisk adferd vil du ha en 90% sannsynlighet for å se en glad deltaker, og en 10% sannsynlighet for å se en bekymret deltaker.
- etter en samtale med en alvorlig adferd er det derimot en 40% sannsynlighet for å se en glad deltaker, og en 60% sannsynlighet for en bekymret deltaker.
- Og etter en samtale med en snakkesalig adferd er det en 70% sannsynlighet for å se en glad deltaker, og en 30% sannsynlighet for en bekymret deltaker.

1) Kan du modellere situasjonen som en Hidden Markov Model hvor robotens adferd er den skjulte tilstanden, og observasjonene tilsvarende de glade eller bekymrede deltakerene som går ut av rommet? Tegn en tabell som beskriver transisjonsmodellen, og en tabell som beskriver emisjonsmodellen.

2) La oss si at 2 deltakere har snakket med roboten etter hverandre. Den første deltakeren forlot rommet med et stort smil, det vil si at *observasjon* _{$t=1$} = "glad". Den andre deltakeren så derimot veldig bekymret ut når hun forlot rommet, det vil si *observasjon* _{$t=2$} = "bekymret". Bruk Viterbi for å beregne sannsynligheten for robotens adferd (altså enten humoristisk, alvorlig eller snakkesalig) i den andre samtalen. Skriv ned alle beregningene dine.

Skriv ditt svar her

Maks poeng: 12

3.1 Dependenssyntaks

Her skal du besvare følgende to spørsmål:

1. Hva kjennetegner såkalt dependensgrammatikk? Forklar kort hvordan dependensgrammatikk skiller seg fra frasestrukturgrammatikk.
2. Det er ikke alltid enkelt å avgjøre hva som er hode og dependent i en setning. Ta utgangspunkt i den vedlagte dependensgrafene og relasjonen mellom *har* og *pekt*. Kan du angi to kriterier for hode-dependentrelasjonen som her står i konflikt? Hvilke er det? Beskriv kort hva konflikten består i og hvilket valg som er tatt her.

Skriv ditt svar her

Maks poeng: 10

3.2 Dependensparsing

Ta utgangspunkt i graf A (vedlagt) og angi en fullstendig transisjonssekvens som vil resultere i denne grafen ved bruk av arc eager-algoritmen for dependensparsing.

Du skal for hver transisjon angi innholdet i stack og inputbuffer samt valgt transisjon og eventuell relasjon som legges til grafen. Merk at du her skal starte og slutte med *kun* rot-etiketten på stacken.

Skriv ditt svar her

Maks poeng: 10

4.1 Maskinoversettelse

1) Hva er et pro-drop eller nullanaforaspråk, og hvorfor kan disse språkene skape problemer for maskinoversettelsesystemer? Forklar med 2-3 setninger.

2) Beregn BLEU for dette (allerede tokeniserte) mini-korpuset som består av to setninger:

Setning 1:

Fasit: Voici un exemple de phrase en francais .

Systemoutput: Voilà un exemple de phrase française .

Setning 2:

Fasit: Le cours était très intéressant .

Systemoutput: Ce cours était très intéressant !

Husk BLEU-formelen: $BLEU = brevity_penalty * (\prod_{i=1}^4 precision_i)^{1/4}$,
hvor $brevity_penalty = \min(1, \frac{\text{antall ord i systemets setninger}}{\text{antall ord i fasitens setninger}})$.

Merk: Tegnsetting bør også betraktes som tokens.

Skriv ditt svar her

Maks poeng: 12

4.2 Chatbots

Vi har i kurset sett på 4 ulike tilnærminger for å utvikle chatbots:

- regelbaserte systemer
- IR-baserte systemer
- sequence-to-sequence systemer
- NLU-baserte systemer

1) For hver av disse fire tilnærmingerne, forklar kort:

a) om utviklingen av systemet krever at man samler inn dialogdata

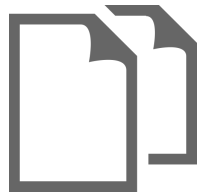
b) hvis ja, om de innsamlede dialogdataene krever noen form for annotering

2) Hva er fordeler og ulemper ved disse 4 tilnærmingerne? Nevn minst én fordel og én ulempe for hver tilnærming.

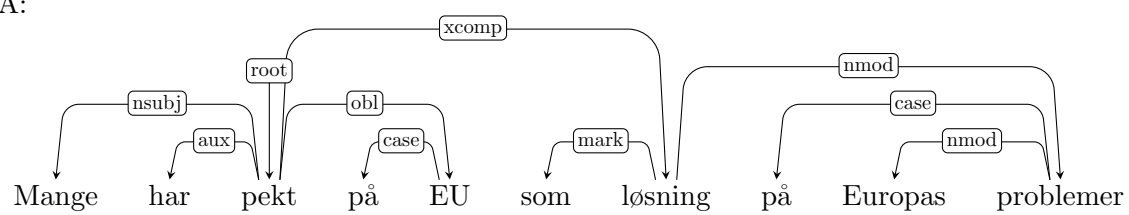
Skriv ditt svar her

Maks poeng: 10

Question 10
Attached



A:



Question 9
Attached



A:

