Løsningsforslag, prøveeksamen IN1160

Vektorrommodeller

1) Vekting med TF-IDF demper vekten til ord som…
   (Poeng: 2 for riktig alternativ; kun ett kan velges)

a. forekommer ofte i ett dokument, men sjelden ellers.
b. forekommer sjelden i både dokumentet og korpuset.
c. forekommer ofte i de fleste dokumenter i korpuset. (RIKTIG)
d. har høy grad av flertydighet.

2) Hvilke av følgende utsagn om cosinus-likhet og euklidisk avstand er sanne når alle vektorer er lengde-normaliserte? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).
   (Poeng: 2 per riktige, -1 per feil, 5 for feilfri besvarelse.)

a. Cosinus-likhet mellom to vektorer er alltid det samme som 1 minus euklidisk avstand delt på 2.
b. Å maksimere cosinus-likhet er ekvivalent med å minimere euklidisk avstand. (RIKTIG)
c. Både cosinus-likhet og euklidisk avstand blir identisk med prikk-produktet og får derfor samme verdi.
d. Hvis cosinus-likhet = 1 så er euklidisk avstand = 0. (RIKTIG)

3) Hva er fordeler ved Rocchio fremfor kNN? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).
   (Poeng: 2 per riktige, -1 per feil, 5 for feilfri besvarelse)

a. Lavere beregnings-kostnad ved klassifikasjon. (RIKTIG)
b. Mere robust mot stor variasjon i størrelse og distribusjon av klasser.
c. Krever ingen lagring av hele treningssettet ved klassifikasjon. (RIKTIG)
d. Håndterer ikke-lineære beslutningsgrenser.
e. Trenger ikke annoterte treningsdata (den er ikke-veiledet).

4) Hvilke påstander om k-Means er riktige? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).
   (Poeng: 2 per riktige, -1 per feil, 7 for feilfri besvarelse)

a. k-Means er en vanlig form for agglomerativ hierarkisk klyngeanalyse.
b. k-Means er en form for veiledet klassifikasjon basert på avstand til de k nærmeste treningseksemplene (`naboene').
c. k-Means er en vanlig form for flat klyngeanalyse. (RIKTIG)
d. k-Means konvergerer alltid til det globale minimumet.
e. k-Means er deterministisk uansett initialisering.
f. k-Means kan konvergere uten nødvendigvis å finne det globale minimumet. (RIKTIG)
g. k-Means krever at parameteret k spesifiseres på forhånd. (RIKTIG)

Klassifikasjon og evaluering

5) Nedenfor ser du noen ulike påstander om såkalt ekstrinsisk og intrinsisk evaluering. Som konkret eksempel på en modell bruker vi tematisk klassifikasjon av dokumenter. Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).
(Poeng: 2 per riktige, -1 per feil, 5 for feilfri besvarelse)

a. Intrinsisk evaluering måler klassifikatorens nytte internt i et større system, f.eks. for personlige innholdsanbefalinger.
b. Ekstrinsisk evaluering måler hvor nyttig klassikatoren er for en større oppgave, f.eks. et nedstrøms system for personlige innholdsanbefalinger. (RIKTIG)
c. Ekstrinsisk evaluering måler hvor nøyaktig modellen ekstraherer informasjon, i dette tilfellet tematiske klasser, sammenliknet med en gullstandard.
d. Mens intrinsisk evaluering måler ytelse på utviklingssdata, bruker ekstrinsisk evaluering testdata (såkalt 'held-out testing').
e. Intrinsisk evaluering måler kvaliteten på selve tekstklassifikasjonen direkte, sammenliknet med en gullstandard. (RIKTIG)
f. Intrinsisk evaluering måler kvaliteten til selve treningsdataene mens ekstrinsisk evaluering måler ytelsen på en trent modell.

Lineær regresjon

6) La oss anta vi har et treningssett med tre punkter, hver med 2 trekk og tilsvarende fasittverdi:

nr	Trekk 1	Trekk 2	Fasittverdi
1	2.0	1.5	5
2	3.5	0.5	4.2
3	1	4	6.3

Vi har estimert en lineær regresjonsmodell basert på dette treningssettet, og har kommet med parameterverdier w_1=-0.25, w_2= 0.42 og b=4.9

Hva er den gjennomsnittlige kvadratfeilen på treningssettet? (NB: bruk kalkulator)

(Poeng: 4 for riktig svar – kun ett riktig):

a) Rundt 0.03
b) Rundt 0.74
c) Rundt 0.0006
d) Rundt 0.001 (RIKTIG)
e) Rundt 0.004

7) Hvilke av følgende utsagn om lineær regresjon er riktige? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).

(Poeng: 2 per riktige, -1 per feil, 5 for feilfri besvarelse)

a) Responsvariabelen i lineær regresjon kan ha negative verdier (RIKTIG)
b) Gradientnedstigning brukes til å måle avviket mellom modellens prediksjoner og fasittverdiene
c) Når alle trekkene er satt til null blir responsen lik konstantleddet b, uansett vektene. (RIKTIG)
d) Hvis vekten $w_i < 0$ og vi øker verdien for trekket x_i vil modellens output y gå opp.
e) En vanlig lineær regresjonsmodell gir oss en sannsynlighetsfordeling over mulige verdier for responsen.

Matriseregning og logistisk regresjon

8) Gitt de to matrisene $A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 4 \end{bmatrix}$ og $B = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, hva er matriseproduktet AB?

(Poeng: 4 for riktig svar – kun ett riktig)

a) $\begin{bmatrix} 5 \\ -1 \end{bmatrix}$ (RIKTIG)
b) $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$
c) $\begin{bmatrix} 3 \\ 7 \end{bmatrix}$
d) $\begin{bmatrix} 5 \\ 15 \end{bmatrix}$
e) Produktet er ikke definert på grunn av inkompatible dimensjoner

9) Hvilke av følgende utsagn om logistisk regresjon er riktige? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).
(2 poeng for riktig svar, -1 per feil. 3 poeng for alt rett og ingen feil)

a) Multinomisk logistisk regresjon bruker sigmoid-funksjonen til å komprimere den vektede summen av inputtrekkene til å være mellom 0 og 1.
b) En logistisk regresjonsmodell med 5 klasser og 8 trekk har til sammen 40 parametere som skal estimeres.
c) I binær logistisk regresjon viser beslutningsgrensen punktene hvor den vektede summen $z = \sum_{I=1}^{p} w_i x_i + b$ er lik null. (RIKTIG)
d) Tapet definert gjennom binær eller kategorisk kryssentropi begrenses til verdier mellom 0 og 1.
e) En logistisk regresjonsmodell gir oss sannsynligheten $P(\mathbf{x} \mid c)$ for å observere inputtrekkene $\mathbf{x}$ når vi vet at punktet kommer fra klassen c.

Nevrale Nett

10) Dette er en oppdateringsregel for vektene/parametrene i en klassifikator:

$$\Delta w_i = \eta(y - \hat{y})x_i$$

Hva representerer η ("eta") i denne regelen?

(Poeng: 2 for riktig alternativ; kun ett kan velges)

a. Gradienten
b. Korrekt klassifikasjon av datapunktet («ground truth»)
c. Læringsraten (RIKTIG)
d. Tapet

11) Hva er sant om algoritmen Tilbakepropagering («backpropagation»)? Kryss av for hvilke påstander som er korrekte (det kan være ett eller flere).

(Poeng: 2 per riktige, -1 per feil, 5 for alle riktige)

a. Den oppdaterer vektene til en modell for å senke tapsverdien på treningsdata. (RIKTIG)
b. Algoritmen er beregningsmessig effektiv fordi den kan hoppe over noen av lagene i det nevrale nettverket
c. Den brukes under trening av ettlags-perseptroner
d. Den brukes når vi tester nevrale nettverk på usette data.
e. Den brukes under trening av nevrale nettverk med ett eller flere skjulte lag (RIKTIG)

Anta vi har et trent flerlags nevralt nettverk for multinomial klassifikasjon av inndata, med 3 mulige klasser. Vi har for eksemplets skyld ikke med noen bias-inndata. Vektverdiene fra det siste skjulte laget til utdata-laget er gitt under. Hvilken klasse vil dette nettverket predikere hvis inndata til det siste laget er **a**=[2,-1,4,0]?

Hint: Rader representerer hvert utdata-nevron, kolonner representerer hvilket nevron disse er koblet til i det siste skjulte laget.

2 1 1

2 2 1

1 2 2

1 1 2

(Poeng: 4 for riktig alternativ; kun ett kan velges)

a. Klasse 3 (RIKTIG)
b. Klasse 1
c. Lik sannsynlighet for alle klasser
d. Klasse 2

Utregning:

Regner ut input-verdi til hvert utdata-nevron:

O1 = (2,2,1,1)*(2,-1,4,0) = 6

O2= (1,2,2,1)*(2,-1,4,0) = 8

O3 = (1,1,2,2)*(2,-1,4,0) = 9

Høyest utverdi tilhører nevronet for klasse 3. Gitt at det ikke er noen tilfeldighet i mappingen mellom ut-verdier og endelig klassifisering, velger nettverket klasse 3. I noen tilfeller vil denne mappingen ha elementer av tilfeldighet, i så fall velger nettverket *mest sannsynlig* klasse 3.

Beslutningstrær og ensembler

12) En «svak lærer» (weak learner) i ensemble-sammenheng er:

(Poeng: 2 for riktig alternativ; kun ett kan velges)

A. En modell som er dårligere enn tilfeldig gjetting
B. En modell som er nøyaktig lik en ekspert
C. En modell som bare så vidt er bedre enn tilfeldig gjetting (RIKTIG)
D. En modell som har null varians

13) Vi har 100 datapunkter: 50 i klasse A og 50 i klasse B.

Entropien E beregnes som

$$E = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

Hvor p_i er andelen i klasse i. Anta $k = 2$. Hva er entropien til dette datasettet? Gi svaret som et desimaltall med én desimal (for eksempel 0.0). (2 poeng)

Riktig svar:

Entropien er
$H = -\big(0.5 \log_2 0.5 + 0.5 \log_2 0.5\big) = 1.0.$

Svar: **1.0**

14) Hva beskriver best en tilfeldig skog (random forest)?

A. Et dypt beslutningstre trent på hele datasettet
B. En metode som trener mange beslutningstrær på tilfeldige underutvalg av datapunkter og/eller trekk (RIKTIG)
C. Et nevralt nettverk med trær som noder

(2 poeng)

ML i praksis

15) Nevn 2 typer statistiske skjevheter som kan oppstå når man samler inn data, og forklar hver type med én setning. *(Poeng: 5)*

- **Utvalgsskjevhet** (*selection bias*): Er dataene representative for alle «enheter» vi ønsker å jobbe med (også kalt *populasjonen* i statistikk)?
- **Måleskjevhet** (*measurement bias*): Er dataene preget av målefeil eller rapporteringsfeil, f.eks. upresise eller inkonsekvente annoteringer?
- **Historisk skjevhet** (*historical / temporal bias*): Er dataene utdaterte eller ikke helt representative for dagens situasjon?
- **Frafallskjevhet** (*attrition / survivorship bias*): Er dataene bevisst eller ubevisst «filtrert», slik at en del av populasjonen falt ut av innsamlingen og mangler i datasettet?
- Og mange andre!

16) Hvilke av de følgende utsagn er riktige? *(2 poeng per riktige svar, -1 per feil, 7 poeng for alt rett)*

a) Når noen datapunkter mangler verdier for noen trekk, bør man erstatte disse med 0.
b) Utliggere bør aldri fjernes fra et datasett
c) Med one-hot encoding erstatter man et kategorisk trekk med en vektor som har samme dimensjon som antall mulige kategorier. (RIKTIG)
d) Etter å ha skalert et numerisk trekk får man et nytt trekk hvor gjennomsnittet og standardavviket er begge = 1.
e) L1 og L2 regularisering gir et fortrinn til modeller som har parametere med lave absoluttverdier. (RIKTIG)
f) Hyperparametere må typisk fastsettes *etter* man har ferdig trent ferdig modellen.
g) Strukturert prediksjon handler om å predikere responser som har en komplisert intern struktur, som en graf, et tre eller en sekvens. (RIKTIG)

17) Vi har et datasett *X* som består av 5 datapunkter og kun ett trekk. Vi ønsker å avdekke mulige *utliggere* i dette datasettet. For å finne disse skal vi ta i bruk en enkel metode, nemlig å klassifisere som utligger alle verdier som befinner seg i *over 2 standardavvik* (pluss eller minus) fra gjennomsnittet for trekket.

Verdiene i *X* er: [0, 3, 5, 9, -2]

Du bør bruke her enten den korrigerte eller ukorrigerte versjonen av standardavviket.

Hvilke verdier vil da bli klassifisert som utliggere ifølge denne metoden? *(Poeng: 4; kun ett svar kan velges)*

a) 9
b) -2
c) 9 og -2
d) Ingen verdi [RIKTIG]

Relevante formler:
For et datasett med verdier $x_1, x_2, \dots, x_n$ og gjennomsnitt $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ så er det ukorrigerte standardavviket gitt ved $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ (brukes typisk når vi ser på datasettet som hele populasjonen) og det korrigerte standardavviket gitt ved $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ (vanligere når datasettet anses som et utvalg fra en større populasjon).

Relevante utregninger underveis:

Gjennomsnitt: $\overline{x} = \frac{0+3+5+9+(-2)}{5} = \frac{15}{5} = 3$

Lista av kvadrerte avstander fra snittet:

$$(0-3)^2 = (-3)^2 = 9$$
$$(3-3)^2 = 0^2 = 0$$
$$(5-3)^2 = 2^2 = 4$$
$$(9-3)^2 = 6^2 = 36$$
$$(-2-3)^2 = (-5)^2 = 25$$

Summen av kvadrerte avvik blir da: $9 + 0 + 4 + 36 + 25 = 74$

Dermed får vi at $\sigma = \sqrt{\frac{74}{5}} = \sqrt{14.8} \approx 3.85$ og $s = \sqrt{\frac{74}{5-1}} = \sqrt{18.5} \approx 4.30$.

Med utgangspunkt i s, som var høyest, får vi da at to standardavvik fra gjennomsnittet bli $3 - 2 \cdot 4.30 \approx -5.60$ og $3 + 2 \cdot 4.30 \approx 11.60$. Vi ser at alle verdiene ligger innenfor intervallet $[-5.60,\ 11.60]$ og riktig svar blir dermed (d) – "ingen verdi".

Forsterkende Læring

18) Hva er **usant** om SARSA?

(ett rett svar, 2p for riktig)

a) SARSA antar vi følger en grådig policy når vi beregner oppdaterte Q-verdier (RIKTIG)
b) SARSA bruker en læringsrate i Q-verdi beregningen
c) SARSA er et eksempel på «på-policy» (on-policy) læring
d) Sarsa oppdaterer iterativt sine estimater av Q

19) Vi har et problem i forsterkende læring der vi skal beregne Q-verdier for par av tilstand/handling. Miljøet har 2 tilstander (S1 og S2) og 2 mulige handlinger i hver tilstand (A1 og A2). Etter å ha utforsket miljøet har vi lært følgende Q-tabell:

Tilstand/Handling	Q-verdi
S1/A1	0.2
S1/A2	0.6
S2/A1	0.0
S2/A2	0.8

Anta at vi er i tilstand S1 og utfører handling A2. Vi får belønning (reward) på 1, og havner i tilstand S2. Beregn oppdatert Q-verdi Q(S1, A2), gitt diskonteringsfaktor lik 0.9 og læringsrate lik 0.1. Bruk den normale oppdateringsfunksjonen fra Q-læring. (svar med 2 desimaler etter komma – 4 poeng)

Utregning:

Husk fra forelesningen at oppdateringsregelen for Q-læring er:

$$Q_{t+1}(s_t, a_t) = \underbrace{Q_t(s_t, a_t)}_{\text{gammel verdi}} + \underbrace{\mu}_{\text{læringsrate}} \times \Bigg[\overbrace{\underbrace{r_{t+1}}_{\text{belønning}} + \underbrace{\gamma}_{\text{diskonteringsfaktor}} \underbrace{\max_a Q_t(s_{t+1}, a)}_{\text{estimat av optimal fremtidig verdi}}}^{\text{lært verdi}} - \underbrace{Q_t(s_t, a_t)}_{\text{gammel verdi}} \Bigg]$$

Her har vi:

- $s_t = S1$, $a_t = A2$
- $Q_t(S1, A2) = 0.6$
- $r_{t+1} = 1$
- $s_{t+1} = S2$
- $\gamma = 0.9$
- $\alpha = 0.1$
- $\max_a Q(S2, a) = \max(0.0, 0.8) = 0.8$

1. Vi finner den lærte verdien:

$$r_{t+1} + \gamma \max_a Q(S2, a) = 1 + 0.9 \cdot 0.8 = 1 + 0.72 = 1.72$$

2. Setter inn i oppdateringsformelen:

$$Q_{t+1}(S1, A2) = 0.6 + 0.1[1.72 - 0.6] = 0.6 + 0.1 \cdot 1.12 = 0.6 + 0.112 = 0.712$$

Svar: Den oppdaterte Q-verdien blir
$Q(S1, A2) = 0.712$.

20) Hva er rollen til diskonteringsfaktoren (discount factor) i forsterkende læring? (4 poeng)

Svar:

Diskonteringsfaktoren γ (mellom 0 og 1) bestemmer hvor mye agenten bryr seg om framtidige belønninger sammenlignet med umiddelbar belønning.

Hvis γ er nær 0, er agenten veldig kortsiktig og bryr seg nesten bare om belønningen nå. Hvis γ er nær 1, er agenten mer langsiktig; framtidige belønninger teller nesten like mye som umiddelbare.

Formelt vekter γ framtidige belønninger i summen av total belønning:

$$R = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots$$

Generativ KI

21) Hva kjennetegner generativ KI sammenlignet med diskriminative modeller?

(ett rett svar, 2p for riktig)

a) Den lærer en beslutningsgrense mellom klasser
b) Den lærer å predikere neste ord uten å bruke data
c) Den lærer mønstre i datadistribusjonen for å skape nye, lignende data (RIKTIG)
d) Den brukes bare til bildeklassifikasjon

22) Hva kjennetegner embeddinger sammenlignet med eldre telle-baserte vektorer?

(ett rett svar, 2p for riktig)

a) De er høy-dimensjonale og svært glisne ("sparse")
b) De er lavdimensjonale, tette og distribuerte (RIKTIG)
c) De inneholder bare binære verdier (0/1)
d) De kan ikke læres med selv-veiledet læring

23) Hva beregnes ved såkalt oppmerksomhet ("attention") i Transformer-arkitekturen?

(ett rett svar, 2p for riktig)

a) en vektet sum av vektorer basert på likhet (RIKTIG)
b) en sannsynlighetsdistribusjon over alle ord i vokabularet
c) modellens tapsfunksjon (ofte ved kryss-entropi)
d) et kvalitetsmål

Filosofi og Fremtid

24) Hva mener vi når vi sier at KI-modeller har «taggete intelligens» (eng.: jagged intelligence)? (5 poeng)

«Taggete intelligens» betyr at KI-modeller kan være svært gode på noen oppgaver og samtidig svært dårlige på andre, nært beslektede oppgaver. Modellen kan løse avanserte, «smarte» oppgaver, men feile på enkle ting som krever robust forståelse, sunn fornuft eller overføring til litt nye situasjoner. Begrepet understreker at KI ikke har en glatt, generell intelligens, men en «taggete» profil med topper (sterke sider) og dype daler (svake sider).

Etikk og samfunnskonsekvenser

25) Du skal vurdere bruken av et automatisk system for sortering av jobbsøknader. Hvilke etiske problemstillinger bør være med i vurderingen? Angi 3 etiske problemstillinger som vil være relevante og beskriv dem kort. (6 poeng)

Her er det en rekke etiske problemstillinger som bør vurderes. Noen eksempler kan være:

- **Persondata** må behandles forsvarlig og i tråd med lovverk. Dersom dataene skal benyttes kreves godkjenning fra personene.
- **Representasjon**: hvem er representert i treningsdata? Underrepresenterte grupper i dataene kan føre til en dårligere ytelse for disse gruppene.
- **Modelltrening**: Bevissthet rundt hva målfunksjonen optimerer. Det kan være faktorer som er relevante for en ansettelse som det ikke optimeres for.
- Historiske **skjevheter** i samfunnet kan forsterkes dersom systemene tas i **bruk** og beslutninger tas med grunnlag i modellene.