



Institutt for datateknikk
og informasjonsvitenskap

Eksamensoppgave i

TDT4171 – Metoder i kunstig intelligens

XX. Aug 2011, kl. 09:00 - 13:00

*Oppgaven er utarbeidet av faglærer Keith Downing og kvalitetssikrer Pauline Haddow.
Kontaktperson under eksamen er Agnar.Aamodt (mobil 93015563)*

Språkform: Bokmål

Tillatte hjelpemidler: D

Ingen trykte eller håndskrevne hjelpemidler tillatt.

Bestemt, enkel kalkulator tillatt.

Sensurfrist: XX

Les oppgaveteksten nøye. Finn ut hva det spørres etter i hver oppgave.

Dersom du mener at opplysninger mangler i en oppgaveformulering gjør kort rede for de antagelser og forutsetninger som du finner det nødvendig å gjøre.

Oppgave 1: Bayesian Networks (35%)

a) Far ser deg (10 %)

Du hadde lovet foreldrene dine å være hjemme innen kl. 23:00, men du glemte tiden og kommer tilbake ved midnatt. Du vet følgende:

1. Ved et tidspunkt under kvelden, er den en sannsynlighet på 0.8 for at lampen på kjøkkenet skrur på.
2. Far er oppe til midnatt eller senere med en sannsynlighet på 0.6.
3. Når far legger seg, sjekker han kjøkkenlampen med en sannsynlighet på 0.7. Når han sjekker det, skrur han det alltid av (sannsynlighet = 1.0) hvis det er på.

Gitt at L = “kjøkkenlampe er på” og F = “far er oppe”, beregn de følgende sannsynligheter:

1. $p(L = \text{sann} :: D = \text{sann})$
2. $p(L = \text{sann} :: D = \text{usann})$
3. $p(L = \text{usann} :: D = \text{sann})$
4. $p(L = \text{usann} :: D = \text{usann})$

der $p(X :: Y)$ angir sannsynligheten for X gitt Y , m.a.o. den betingende sannsynligheten for X når vi vet Y .

Du kan se at kjøkkenlampa er på fra utsiden av huset. Bruk Bayes regel til å beregne sannsynligheten for at far fremdeles er oppe (og venter på deg) gitt at kjøkkenlampa er på.

Vis alt arbeidet ditt tydelig, og lag en “CIRCLE” rundt hver av de 5 sannsynlighetene det spørres etter.

b) Mors lille hjelper (15 %)

Du har nettopp kjøpt en mobiltelefon, men bare EN person kan nummeret dit, nemlig din mor. Betingelsene som ligger til grunn for at hun ringer deg er kompliserte og trenger å kartlegges med et Bayesiansk nettverk. Dette er alt du vet:

1. Hvis det var en SOLRIK dag, vil din mor jobbe i hagen. Sannsynligheten for at hun TRENGER noe (fra din side) er 0.8 når hun gjør dette.
2. Hvis mor ikke går ut i hagen, synker sannsynligheten for at hun trenger noe til 0.1.
3. Sannsynligheten for at mor bare er PRATESJUK på en gitt dag, er 0.5.
4. Sannsynligheten for at mor vil RINGE når hun TRENGER noe er 0.6.
5. Sannsynligheten for at mor vil RINGE deg når hun er PRATESJUK er 0.9.
6. Allikevel, hvis hun både er PRATESJUK og TRENGER noe, er sannsynligheten for at hun RINGER bare 0.3, fordi hun ofte vil skravle med naboen og glemme det hun trenger.
7. Hvis mor verken TRENGER noe eller er PRATESJUK, vil hun ikke RINGE deg.
8. Sannsynlighetene for at en dag er SOLRIK er 0.7.

9. Hvis mor prøver å RINGE deg, er sannsynligheten for at telefonen RINGER 0.95, fordi teleselskapet ikke er helt pålitelig.
10. Allikevel, hvis du ikke blir oppringt av noen (altså din mor), kan teleselskapet RINGE deg direkte (for å oppfordre deg til å bytte til et mer lønnsomt abonnement), med en sannsynlighet på 0.02.

Sett opp et fullstendig Bayesiansk nettverk som bruker de 5 kaptelérte uttrykkene over (som dine variabler). Gjør fornuftige valg for de uavhengige og avhengige variablene, sammen med foreldre-barn-relasjoner. Bruk alle kjente verdier (basert på informasjonen over) i sannsynlighetstabellene.

c) Uavhengighet (5 %)

På spørsmålene under er A, B og C hendelser, og $P(X | Y)$ angir den betingende sannsynligheten for X gitt Y.

1. Hvis $P(A) = P(B)$, betyr dette at A og B er uavhengige?
2. Hvis $P(A :: B) = P(B)$, betyr dette at A er betinget uavhengig av B?
3. Hvis $P(A :: B \text{ og } C) = P(A :: C)$, betyr dette at A er betinget uavhengig av B, gitt C?
4. Hvis $P(A \text{ og } B :: C) = P(A :: C) \times P(B :: C)$, betyr dette at A er betinget uavhengig av B, gitt C?
5. Når du beregner den unionsfordelingstabellen (joint distribution table) for A, B og C, hjelper det å vite at A og B er betinget uavhengige i forhold til å redusere det nødvendige antallet påkrevde a-priori sannsynligheter med minst 50%?

d) Filtrere, glatte, forutse (5 %)

For hvert av spørsmålene under, vis ganske enkelt hvorvidt det er et eksempel på å forutse (P), glatting (G) eller filtrering (F). Ingen ytterligere forklaring er nødvendig.

1. Basert på alle kredittkorttransaksjoner før og etter ranet, jobbet bandittene for mafiaen da de ranet Chase Manhattan Bank?
2. Ut fra det vi nå vet om disse bandittene, vil de rane en ny bank i 2011?
3. Gitt alt du har sett, Dr. Jekyll, lider denne pasienten av schizofreni nå?
4. Basert på et mønster dannet av blodflekker i offerets oppkjørsel, rømte morderen til fots eller i bil?
5. Basert på all tilgjengelig bevismateriale fra fossiler og geologi, begynte en istid for 500 millioner år siden?

Oppgave 2: Beslutningsnettverk (20%)

- a) Hvilke tre typer noder består et beslutningsnettverk (decision network, influence diagram) av? Og hva representerer de?
- b) En person ønsker å kjøpe en bruktbil. Han kan gjøre noen tester av bilen, og så beslutte om han vil kjøpe eller ikke. Bilen kan være av god eller dårlig kvalitet, og resultatet av testen kan si noe om kvaliteten.
Tegn et beslutningsnettverk som representerer dette problemet. Presiser og begrunn alle ytterligere antakelser du gjør.

Oppgave 3: Case-basert resonnering (15%)

- a) Case-basert resonnering kan ses på som en metode for integrert problemløsning og læring. Forklar hvorfor. Bruk gjerne CBR-syklusen (The CBR cycle) som referanse.
- b) Case-baserte læringsmetoder karakteriseres ofte som "lat læring" (lazy learning), i motsetning til andre læringsmetoder som kalles "ivrige" (eager learning). Kan du tenke deg hvorfor?

Oppgave 4: Læring og Beslutningstrær (15%)

a) Læring

1. Hvorfor skulle en agent ønske å lære?
2. Nevn alle de tre hovedtypene tilbakemelding som man kan lære fra, og beskriv dem kort.
3. Hvilke kriterier er relevante når man skal velge læringsmetode? Lag en liste av dem og beskriv dem kort.

b) Beslutningstrær

1. Forklar et beslutningstres syntaks.
2. Gi en uformell beskrivelse av læringsalgoritmen ID3 for et beslutningstre.
3. Hva er «overfitting» (overtilpasning)?
4. Gi et enkelt eksempel på et beslutningstre med minst fire attributter, hvorav minst en ikke er boolsk. Vis minst fire læringseksempler i tabellformat, og illustrer eksemplene i et beslutningstre. Forklar hvordan du kom frem til treet fra tabellformatet.

Oppgave 5: Språkprosessering (10%)

La os si du har tilgang til de følgende kunnskapskilder:

- ordbøker over grunnleggende ordformer (lemma) for engelsk og norsk
- morfologiske bøyningsegler for de samme to språk
- en stor mengde engelsk e-post som allerede er klassifisert som spam
- en stor mengde norsk e-post som allerede er klassifisert som spam
- en stor mengde engelsk e-post som allerede er klassifisert som ikke-spam
- en stor mengde norsk e-post som allerede er klassifisert som ikke-spam
- en enorm mengde uklassifisert e-post som skrevet på et vidt spekter av menneskelige språk
- en strøm av inngående twitter-meldinger (kvitringer), hvorav ingen er lengre enn en setning

Beskriv hvordan du vil bygge et system som kan analysere hver inngående twitter-melding og produsere en av de følgende outputene:

1. meldingen er ikke skrevet på et menneskelig språk
2. meldingen er skrevet på et menneskelig språk som verken er engelsk eller norsk
3. meldingen er skrevet på engelsk og er spam
4. meldingen er skrevet på norsk og er spam
5. meldingen er skrevet på engelsk og er ikke spam
6. meldingen er skrevet på norsk og er ikke spam

Du trenger ikke å gi en fullstendig løsning, men heller å skissere de nødvendige skrittene for å nå dit.

Oppgave 6: Bildeprosessering (5%)

Tegn et annotert diagram som viser de skrittene som må tas for å gjøre automatisk gjenkjenning av et bilde, eksempelvis i en inspeksjon i industrien.

Hvilket av disse skrittene er ansett som det vanskeligste å få til å fungerer ordentlig for å understøtte systemets formål?