

**Norges teknisk-naturvitenskapelige universitet
Institutt for datateknikk og informasjonsvitenskap**

<i>EKSAMEN i fag</i>	TDT4171 Maskinl�ring og case-basert resonnering
<i>Dato</i>	15. desember 2008
<i>Antall timer</i>	4
<i>Antall studiepoeng</i>	7,5
<i>Antall sider</i>	2
<i>Tillatte hjelpemidler</i>	Ingen
<i>Sensurdato</i>	12. januar 2009

Alle 4 oppgavene, med alle deloppgaver, skal besvares. Hver del-oppgave tillegges like stor vekt ved bedømmelsen.

Faglig kontakt under eksamen: Professor Agnar Aamodt, Institutt for datateknikk og informasjonsvitenskap, tlf. 735 91838

Oppgave 1 – Basisbegreper

- En fundamental antakelse for all l ring er uttrykt i *hypotesen for induktiv l ring* (The Inductive Learning Hypothesis). Hva uttrykker denne hypotesen?
- En eller annen form for bias er n dvendig for l ring.
 - Hvorfor?
 - Gi et eksempel p  en metode uten bias.
- Gitt de to formene for bias, hhv. preference bias (search bias) og restriction bias (language bias).
 - Hva er Naive Bayes restriction bias?
 - Hva er search bias for Support Vector Machines?
- Hva menes med *overtilpasning (overfitting)* i sammenheng med induktiv l ring? Hva kan for rsake overtilpasning, og hvordan kan man unng  overtilpasning? Du kan eksemplifisere ved   referere til l ring av beslutningstr er dersom  nskelig.

Oppgave 2 - Casebasert resonnering

- Nevn den prinsipielle forskjellen mellom k-Nearest Neighbour metoden og case-basert resonnering i mer generell forstand, for hvert av de f lgende to trinn i CBR-syklusen:
 - Retrieve
 - Reuse

- b) Gitt CBR-systemene Protos og Creek. Beskriv for hver av de to systemene:
- I hvilke av CBR-syklusens faser benyttes generell domenekunnskap?
 - Hvordan benyttes den generelle domenekunnskapen i Retrieve fasen?
- c) I artikkelen "Remembering to Forget" foreslås det en modell for å bestemme en case-bases kompetanse. Hva uttrykker begrepene
- "retrieval space"
 - "adaptation space"
 - "coverage set"
 - "reachability set"?

Skisser kort hvordan dette rammeverket brukes til å definere en strategi for å slette case fra en case-base.

Oppgave 3 - Bayesiansk læring

- a) Hva karakteriserer en Naive Bayes klassifikator? Nevn dens styrker og svakheter, og angi spesielt antagelsen en Naive Bayes klassifikator gjør om (betingede) uavhengigheter mellom de forskjellige variable. Baserer Bayesianske nett seg på de samme antagelsene som Naive Bayes? Begrunn svaret.
- b) EM algoritmen består av to trinn ("E-trinnet" og "M-trinnet") som gjentas iterativt. Beskriv med egne ord hva som skjer i de to trinnene.
- c) Noen maskinlæringalgoritmer utnytter apriori kunnskap, andre ikke. Gi minst ett eksempel på typer av læremetoder (fra pensum) som benytter apriori kunnskap, og ett som ikke gjør det. For de(n) metodene som gjør det, beskriv kort hvilken rolle apriori-kunnskapen spiller i læremetoden.

Oppgave 4 – Blandete spørsmål

- a) Hvordan brukes "ensemble metoder" til å forbedre klassifikasjons-algoritmer svake ("weak learners")? Beskriv forskjeller og likheter mellom *boosting* og *bagging*.
- b) Hva er hovedideen i "sekvensiell dekning" (sequential covering) algoritmen? I hvilke situasjoner vil det være lurt å bruke denne algoritmen? Begrunn svaret.