

Eksamensspørgsmål

DAT8, ML-delen af DWML

Forår 2002

Nedenstående udgør eksamensspørgsmålene i ML-delen af kurset “Data Warehousing og Machine Learning” (DWML).

Den relevante litteratur i forbindelse med nedenstående eksamensspørgsmål udgøres af følgende afsnit i Tom Mitchell’s bog “Machine Learning”:

Concept Learning: Hele kapitel 2

Decision Trees: Hele kapitel 3

Neural Networks: 4.1, 4.2, 4.3, 4.4, 4.5, 4.6

Bayesian Learning: 6.1, 6.2, 6.3, 6.7, 6.8, 6.9, 6.10

I nedenstående refererer ‘ML’ til ovennævnte bog.

1 Niveau

Kursets formål er, at den enkelte studerende opnår viden om gængse metoder og algoritmer til automatisk maskinel indlæring på baggrund af data, herunder metoder og algoritmer til indlæring af hypoteser, som kan forklare givne fænomener (“target concepts”).

Da dette emneområde er temmelig omfattende og ikke med rimelighed kan dækkes i et kursus med et omfang på kun 1.5 ECTS, har det ikke været hensigten, at den studerende skulle opnå detaljeret kendskab til de behandlede emner, herunder metoder og algoritmer. Det har derimod været hensigten at den studerende opnåede kendskab til området på et mere overordnet niveau, og specielt blev i stand til at vurdere de enkelte metoder og algoritmers anvendelighed til løsning af problemer under forskellige forudsætninger.

Derfor forventes den studerende til eksamen at kunne demonstrere overblik snarere end kendskab til detaljer vedrørende de i kurset berørte metoder og algoritmer. For at demonstrere et sådant overblik, forventes den enkelte studerende, til hver af de berørte metoder og algoritmer, at kunne redegøre for en række generelle spørgsmål. Herunder hører f.eks. følgende spørgsmål:

- *When*: Hvornår er denne metode velegnet?
- *Hypothesis space*: Hvordan beskrives hypotese-rummet for denne metode, og hvordan repræsenteres hypoteserne?
- *Robustness*: Er denne metode robust overfor støjfyldte data?
- *Incomplete data*: Kan denne metode håndtere ufuldstændige cases (dvs. cases, hvor én eller flere attributværdier mangler)?

2 Eksamensspørgsmål

Nedenfor følger en liste af eksamensspørgsmål, hvor der til hvert spørgsmål er angivet en række underemner (inkluderende ovenstående generelle spørgsmål, hvor disse er relevante), som udgør de begreber/spørgsmål, der anses for essentielle. Emner berørt i bogen (ML), som ikke indgår i nedenstående anses for mindre væsentlige, specielle og/eller mere avancerede end eksaminanden forventes at kunne redegøre fyldestgørende for.

Emnerne og de tilhørende underemner er angivet på engelsk af hensyn til konsistens med den i bogen (ML) anvendte terminologi.

1. Concept Learning

- target function
- training examples
- hypothesis space
- inductive learning hypothesis
- general-to-specific ordering
- FIND-S
- version space
- CANDIDATE-ELIMINATION
- inductive bias
- futility of unbiased learning

2. Decision Tree Learning

- definition (representation and representational power)
- hypothesis space
- when to consider decision-tree learning

- construction algorithm (ID3)
- attribute selection (information gain)
- inductive bias, Occam's razor
- robustness to noisy data
- overfitting and pruning
- decision trees vs. sets of rules
- ability to handle incomplete data

3. Neural Networks

- definition (representation)
- hypothesis space
- when to consider neural networks
- perceptron
- linear separability, representational power
- perceptron training
- gradient descent and delta rule
- incremental gradient descent
- sigmoid threshold unit (why?)
- backpropagation
- convergence and local minima
- expressive capability (representational power)
- inductive bias
- why hidden layer?
- overfitting, stopping criterion

4. Bayesian Learning

- Bayes' theorem and other basic formulas
- prior and posterior probability
- MAP and ML hypothesis
- hypothesis space
- relation to concept learning
- Bayes optimal classifier
- naive Bayes classifier
- m-estimate (equivalent sample size)

3 Betingelser

Eksaminanden trækker ét af ovenstående spørgsmål, og får derefter — uden forberedelsestid — 20 minutter til at redegøre for det til spørgsmålet hørende emne. Der gives karakter efter 13-skalaen.

Uffe Kjærulff, 6. juni 2002

Held og Lykke!!