

# Tentamen i Maskininlärning 2D1431

2004-12-15, kl 14.00 – 19.00

Tillåtna hjälpmedel: *Miniräknare*.  
Preliminära betygsgränser:

$$\begin{aligned} 0 \leq p < 23 &\rightarrow U \\ 23 \leq p < 29 &\rightarrow 3 \\ 29 \leq p < 35 &\rightarrow 4 \\ 35 \leq p \leq 40 &\rightarrow 5 \end{aligned}$$

## 1 Version Space

Beskriv hur *hypotesrummet* och *version space* hänger samman. Vilken sorts element har dessa mängder och vilken är normalt störst? Är det alltid så?

(2p)

## 2 Överträning

Överträning är ett problem som kan drabba många inlärningsalgoritmer. Under vilka förhållanden är det störst risk att råka ut för detta? Hur bör man göra för att upptäcka problemet?

(2p)

## 3 Enlagers neuronät

Bestäm vikterna i ett enlagers trösklat neuronät med två ingångar så att det beräknar funktionen  $a \wedge \neg b$ .

(3p)

Sanningstabellen skall alltså vara

$a$	$b$	Resultat
<b>False</b>	<b>False</b>	<b>False</b>
<b>False</b>	<b>True</b>	<b>True</b>
<b>True</b>	<b>False</b>	<b>False</b>
<b>True</b>	<b>True</b>	<b>False</b>

Du måste förklara (gärna med en figur) vilka vikter du refererar till. Ange också hur du väljer att representera **True** och **False**.

## 4 Bias

(3p)

- Vad menas med begreppen *restriction bias* och *preference bias*?
- Ge exempel på en konkret inlärningsalgoritm som uppvisar *restriction bias*. Ange vilken biasen är.
- Ge exempel på en konkret inlärningsalgoritm som uppvisar *preference bias*. Ange vilken bias man får här.

## 5 VC-dimension

(3p)

Betrakta en hypotesmängd  $H$  bestående av alla *axelparallella rektanglar i planet*. Hypoteserna är alltså tupler av fyra tal  $\langle a, b, c, d \rangle$  som betyder

$$(a < x < b) \wedge (c < y < d)$$

medan exemplen är talpar  $\langle x, y \rangle$ .

Vad är VC-dimensionen för  $H$ ? Bara rätt svar ger inga poäng, du måste visa hur du kommer fram till svaret.

## 6 Entropi

(3p)

Givet data från tabellen nedan som beskriver hur exempel bestående av fyra attribut  $(a_1, \dots, a_4)$  klassas ( $c$ ).

- Beräkna *entropin* för exempelmängden.
- Hur stor blir den förväntade *informationsvinsten* ifall man delar upp datamängden m.a.p. attribut  $a_1$ ?

$a_1$	$a_2$	$a_3$	$a_4$	$c$
0	0	0	0	-
0	0	1	0	+
1	1	0	0	-
0	1	1	1	+
1	0	1	1	+
0	0	1	1	-

För full poäng krävs att du visar vilka formler du använt för att räkna fram svaren.

## 7 Tärningsspel

(6p)

Ett tärningsspel har följande regler:

- Du slår med en tärning upprepade gånger och får själv välja när du ska sluta.
  - Du summerar poängen från alla tärningsslag.
  - Om sammanlagda poängen är högst 12 när du slutar så vinner du motsvarande antal kronor. Om poängen är över 12 så vinner du ingenting.
- Betraktat som ett belöningsbaserat inlärningsproblem, vad utgör här *handlingar* (actions) och *tillstånd* (states)?
  - Räkna ut den *optimala policyn*.
  - Vad är den förväntade vinsten när man börjar spela?

## 8 Julklappar

(4p)

Kalle har ovanligt nog gjort iordning alla julklappar i god tid i år. Redan i somras virkade han 20 grytlappar och köpte 10 ultravåldsamma datorspel som han slog in i 30 prydliga paket. Tyvärr glömde han att skriva upp vad han lade i de olika paketen. Han har dock en svag minnesbild av att ungefär hälften av grytlapparna blev inslagna i guldpaper medan resten fick rött paper. När det gäller datorspelen minns han bestämt att guldpaperet bara räckte till tre, medan resten slogs in i rött paper.

När det är dags att ge mormor sitt paket plockar Kalle på måfå ett ur högen. Det råkar bli ett rött paket.

Vad innehåller mormors paket troligen enligt ett MAXIMUM-LIKELIHOOD resonemang? Vad bör innehållet vara enligt ett MAXIMUM-A-POSTERIORI (MAP) resonemang?

Du måste visa hur du gör beräkningarna enligt de två metoderna!

## 9 PAC-learning

(2p)

PAC är en förkortning för "Probably Approximately Correct". Vad menas här med "Approximately"; vad menas med "Probably". Du ska alltså inte förklara vad orden betyder i vardagsspråk, utan vad de har för precis betydelse inom inlärningsteorin.

Vad krävs för att man ska få påstå att ett problem är *PAC-learnable*?

## 10 Genetiska algoritmer

(6p)

Du vill använda en *genetisk algoritm* för att bestämma vikterna i ett tvålagers framåtkopplat neuronät. Du har ett hundratal tränings exempel på formen  $x, y, z \mapsto r$  där alla värden  $x, y, z$  och  $r$  är reella tal i intervallet  $[-1, 1]$ .

- Vad utgör *individerna* i den genetiska algoritmen i detta fall? Vad är en lämplig *kodning*?
- Vad bör du använda som *fitness-funktion*?
- Beskriv minst ett sätt att göra korsningar (*crossover*) mellan individerna i detta exempel.

## 11 k-Nearest Neighbor

(3p)

Använd 3-NEAREST-NEIGHBOR för att klassificera mönstret  $\langle 0, 1, 0, 1 \rangle$  när man har följande tränings exempel:

$\langle 0, 0, 0, 0 \rangle$	$\mapsto$	Negativt
$\langle 0, 1, 0, 0 \rangle$	$\mapsto$	Positivt
$\langle 0, 0, 1, 0 \rangle$	$\mapsto$	Negativt
$\langle 0, 0, 0, 1 \rangle$	$\mapsto$	Positivt
$\langle 1, 0, 0, 0 \rangle$	$\mapsto$	Negativt
$\langle 0, 0, 1, 1 \rangle$	$\mapsto$	Positivt
$\langle 0, 1, 0, 1 \rangle$	$\mapsto$	Negativt
$\langle 1, 0, 1, 0 \rangle$	$\mapsto$	Positivt
$\langle 1, 1, 1, 0 \rangle$	$\mapsto$	Negativt
$\langle 1, 1, 1, 1 \rangle$	$\mapsto$	Positivt

Du måste naturligtvis redovisa *hur* du kommit fram till svaret för att få poäng på uppgiften.

## 12 Beslutsträd

(3p)

Konstruera ett beslutsträd som motsvarar följande predikat:

$$(a \wedge \neg b \wedge c) \vee (b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge \neg c)$$

Variablerna  $a$ ,  $b$  och  $c$  står för attribut som kan vara sanna eller falska i exemplen.

# Exam in Machine Learning 2D1431

2004-12-15, kl 14.00 – 19.00

Aids allowed: *calculator*.

Preliminary grades:

$0 \leq p < 23$	$\rightarrow$	$U$
$23 \leq p < 29$	$\rightarrow$	3
$29 \leq p < 35$	$\rightarrow$	4
$35 \leq p \leq 40$	$\rightarrow$	5

## 1 Version Space

(2p)

Describe how the *hypothesis space* and the *version space* are related. What constitutes elements in these sets and which set is normally largest? Is this always true?

## 2 Overtraining

(2p)

Overtraining is a problem for many learning algorithms. Under what conditions is the risk largest for overtraining? What can be done to detect the problem?

## 3 Single layer neural network

(3p)

Calculate the weights of a single layer neural network with two inputs, så that it implements the functions  $a \wedge \neg b$ .

The truth table for this function is

$a$	$b$	Result
<b>False</b>	<b>False</b>	<b>False</b>
<b>False</b>	<b>True</b>	<b>True</b>
<b>True</b>	<b>False</b>	<b>False</b>
<b>True</b>	<b>True</b>	<b>False</b>

You must explain (preferably with a figure) what weights you are referring to. Also state how you choose to represent **True** and **False**.

## 4 Bias

(3p)

- What do the terms *restriction bias* and *preference bias* mean?
- Give one example of a concrete learning algorithm with *restriction bias*. State what bias this algorithm has.
- Give one example of a concrete learning algorithm with *preference bias*. What is the bias here?

## 5 VC-dimension

(3p)

Consider a set of hypotheses,  $H$ , consisting of all *rectangles in the plane, parallel to the coordinate axes*. Thus, the hypotheses are tuples of four numbers  $\langle a, b, c, d \rangle$ , meaning

$$(a < x < b) \wedge (c < y < d)$$

while the examples are pairs of numbers  $\langle x, y \rangle$ .

What is the VC-dimension of  $H$ ? Only the correct answer does not give any points; you must show how you have reached the answer.

## 6 Entropy

(3p)

The data in the table below describes the classification  $c$  of different examples consisting of four attributes  $(a_1, \dots, a_4)$ .

- Calculate the *entropy* for this set of samples.
- How large is the expected *information gain* when the set is partitioned using attribute  $a_1$ ?

$a_1$	$a_2$	$a_3$	$a_4$	$c$
0	0	0	0	-
0	0	1	0	+
1	1	0	0	-
0	1	1	1	+
1	0	1	1	+
0	0	1	1	-

To get full points you have to show which formulae you use to get the answers.

## 7 Playing with a dice

(6p)

A dice-game has the following rules:

- You roll a dice as many times as you want; you decide when to stop.
  - Each time you roll the dice, the points are added to the total sum.
  - If the final total sum is at most 12 when you stop, then you win the same amount of kronor. If the sum is over 12, then you don't get anything.
- Regarded as a reinforcement learning problem, what constitutes the *actions* and the *states* here?
  - Calculate the *optimal policy*.
  - What is the expected gain when you start the game?

## 8 Christmas gifts

(4p)

Charley did prepare all his Christmas gifts well in advance this year. Already during the summer, he did crochet 20 potholders and he also bought 10 ultraviolet computer games. He wrapped all these things into 30 neat packages. Unfortunately, he forgot to note what he put in each package. He has a diffuse memory of about half of the potholders being wrapped in gold-paper and the rest in red paper. For the computer games, he definitely remembers that he only had gold-paper for three of them, while the rest were wrapped in red.

When the time has come for Charley to give a Christmas gift to his grandmother, he picks one at random from his bag. It happens to be a red package.

What is the most likely contents of grandmothers package, according to a MAXIMUM-LIKELIHOOD reasoning? What's the contents according to a MAXIMUM-A-POSTERIORI (MAP) reasoning?

You must show how you make the calculations according to the two methods!

## 9 PAC-learning

(2p)

PAC is an abbreviation for "Probably Approximately Correct". What does the term "Approximately" mean here? What does the term "Probably" mean? Note that you are not supposed to explain the ordinary meaning of these words, but their precise meaning in learning theory.

What is required to say that a problem is *PAC-learnable*?

## 10 Genetic algorithms

(6p)

You intend to use a *genetic algorithm* to set the weights of a two-layered feedforward neural network. You have about one hundred training examples of the form  $x, y, z \mapsto r$  where all values  $x, y, z$ , and  $r$  are real numbers in the interval  $[-1, 1]$ .

- a) What constitutes the *individuals* in the genetic algorithm in this case? What is a suitable *coding*?
- b) What should be used as the *fitness-funktion*?
- c) Describe at least one way how to make *crossovers* between the individuals in this example.

## 11 k-Nearest Neighbor

(3p)

Use 3-NEAREST-NEIGHBOR to classify the pattern  $\langle 0, 1, 0, 1 \rangle$  given the following training patterns:

$\langle 0, 0, 0, 0 \rangle$	$\mapsto$	Negative
$\langle 0, 1, 0, 0 \rangle$	$\mapsto$	Positive
$\langle 0, 0, 1, 0 \rangle$	$\mapsto$	Negative
$\langle 0, 0, 0, 1 \rangle$	$\mapsto$	Positive
$\langle 1, 0, 0, 0 \rangle$	$\mapsto$	Negative
$\langle 0, 0, 1, 1 \rangle$	$\mapsto$	Positive
$\langle 0, 1, 0, 1 \rangle$	$\mapsto$	Negative
$\langle 1, 0, 1, 0 \rangle$	$\mapsto$	Positive
$\langle 1, 1, 1, 0 \rangle$	$\mapsto$	Negative
$\langle 1, 1, 1, 1 \rangle$	$\mapsto$	Positive

Of course, you must show *how* you have reached your answer in order to get any points on this problem.

## 12 Decision Trees

(3p)

Construct a decision tree corresponding to this predicate:

$$(a \wedge \neg b \wedge c) \vee (b \wedge \neg c) \vee (\neg a \wedge \neg b \wedge \neg c)$$

The variables  $a$ ,  $b$  och  $c$  denote attributes which can be true or false in the examples.