Christoph F. Eick

Homework2 COSC 6368 Fall 2016

Second Draft

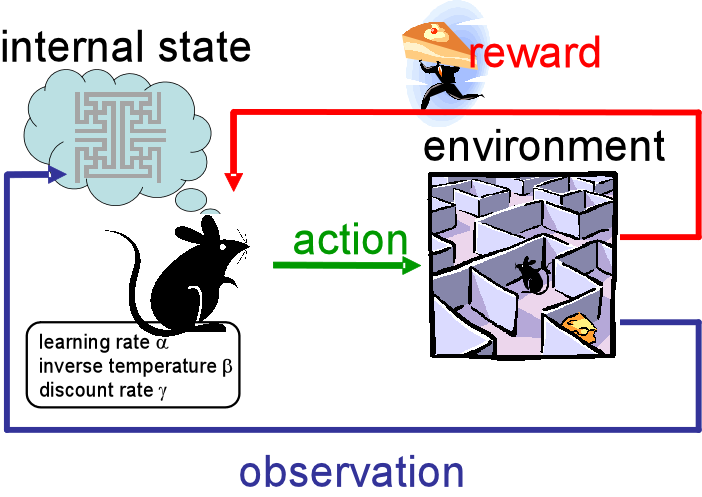
Individual Homework[[1]](#footnote-1) except for the Last Problem[[2]](#footnote-2)

Deadlines: Wednesday, November 30, 11p (problems 1-5); Saturday, December 3, 11p (last problem!)—*late submissions will not be graded*!

Weights: Problems 1-5 count 4% towards the overall course grade; problem 6 counts 2% towards the overall grade.

Last updated: Nov. 16, 2016, 10p

1. Reinforcement Learning w=2



a) Compute the Bellman equations different states in the STU World (<http://www2.cs.uh.edu/~ceick/ai/STU-World.ppt>)!

b) Consider temporal difference learning and Bellman update; in which situations would you use which approach? [3]

c) What does the discount factor γ measure?

2) Information Gain, Entropy and Learning from Examples in General w=3

a) Assume we have a classification problem involving 3 classes: professors, students, and staff members. There are 750 students, 150 staff members and 100 professors. All professors have blond hair, 50 staff members have blond hair, and 250 students have blond hair. Compute the information gain of the test *“hair-color=’blond’”* that returns true or false. Give the formula you used to compute the information gain as well as the actual value! Use H as the entropy function in your formula (e.g. H(1/3,1/6,1/2) is the entropy that 1/3 of the examples belong to class1 1/6 of the examples belong to class 2, and half of the examples belong to class 3).

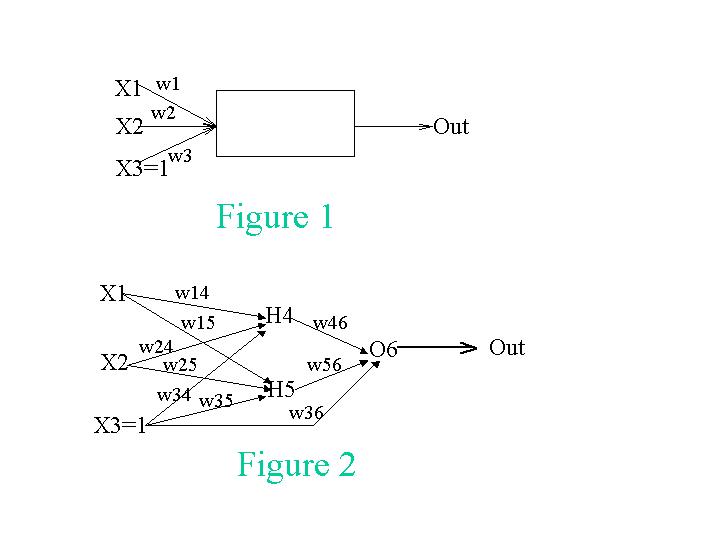
b) What is the “intuitive idea” that underlies ID3’s information gain heuristic --- what attributes does it prefer?

c) Assume you have a training set that does not contain any inconsistent examples and you apply the decision tree induction algorithm to that training set; what is the maximum training accuracy you can achieve? Give reasons for your answer!

d) Assume you have learnt a decision tree from a training set and you observe overfitting; what measures could you take to reduce overfitting?

3) Neural Networks w=6

Assume we have the perceptron that is depicted in Fig. 1 that has two regular inputs, X1 and X2, and an extra fixed input X3, which always has the value 1. The perceptron's output is given as the function:   
  
Out= If (w1\*X1 + w2\*X2 + w3\*X3) > 0 then 1 else 0  
  
Note that using the extra input, X3, we can achieve the same effect as changing the perceptron's threshold by changing w3. Thus, we can use the same simple perceptron learning rule presented in our textbook to control this threshold as well.



**A.** We want to teach the perceptron to recognize the function X1 XOR X2 with the following training set:

|  |  |  |  |
| --- | --- | --- | --- |
| **X1** | **X2** | **X3** | **Out** |
| **1** | **1** | **1** | **0** |
| **0** | **1** | **1** | **1** |
| **1** | **0** | **1** | **1** |
| **0** | **0** | **1** | **0** |

Show the change in the weights of the perceptron for every presentation of a training instance. Assume the initial weights are: w1=0.1, w2=0.1, w3=0.9 Important: Do the iterations according to the order of the samples in the training set. When you finish the four samples go over them again. You should stop the iterations once you get convergence, or when the values you get indicate that there will be no convergence. In either case explain your decision to stop the iterations. Assume in your computations that the learning rate  is 0.4.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample#** | **X1** | **X2** | **X3** | **Output** | **True\_Out** | **Error** | **w1** | **w2** | **w3** |
| **0** |  |  |  |  |  |  | **0.1** | **0.1** | **0.9** |
| **1** | **1** | **1** | **1** |  | **0** |  |  |  |  |
| **2** | **0** | **1** | **1** |  | **1** |  |  |  |  |
| **3** | **1** | **0** | **1** |  | **1** |  |  |  |  |
| **4** | **0** | **0** | **1** |  | **0** |  |  |  |  |
| **5** | **1** | **1** | **1** |  | **0** |  |  |  |  |
| **6** | **...** | **...** | **...** | **...** | **...** | **...** | **...** | **...** |  |
| **7** |  |  |  |  |  |  |  |  |  |
| **8** |  |  |  |  |  |  |  |  |  |
| **...** |  |  |  |  |  |  |  |  |  |

**B.** This time, instead of being limited to a single perceptron, we will introduce hidden units and use a different activation function. Our new network is depicted in Fig. 2. Assume that the initial weights are w14 = 0.1, w15 = 0.1, w24 = 0.1, w25 = 0.1, w34 = 0.1, w35 = 0.1, w36 = 0.3, w46 = 0.3, and w56 = 0.3. The training set is the same as in (A). Use =0.3 as your learning rate. Show what the new weights would be after using the backpropagation algorithm for two updates using just the first two training instances. Use g(x) = 1/(1+e\*\*(-x)) as the activation function; that is g'(x)=(e\*\*(-x))/(1+e\*\*(-x))\*\*2).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S#** | **X1** | **X2** | **X3** | **Out** | **True\_Out** | **Error** | **w14** | **w15** | **w24** | **w25** | **w34** | **w35** | **w36** | **w46** | **w56** |
| **0** |  |  |  |  |  |  | **0.1** | **0.1** | **0.1** | **0.1** | **0.1** | **0.1** | **0.3** | **0.3** | **0.3** |
| **1** | **1** | **1** | **1** |  | **0** |  |  |  |  |  |  |  |  |  |  |
| **2** | **0** | **1** | **1** |  | **1** |  |  |  |  |  |  |  |  |  |  |

4. Bayes’ Theorem (w=2)



Thomas Bayes ≈1740

1. Assume we have 3 symptoms S1, S2, S3 and the following probabilities: P(D)=0.02 P(S1)=P(S2)=P(S3)=0.01; P(S1|D)=0.1; P(S2|D)=0.02; P(S3|D)=0.002. How would a naïve Bayesian system compute the following probability [2]?

P(D|S1,S2,S3)=…

b) Now assume the following additional knowledge has become available: P(S1,S2)=0.0002; P(S3|S1,S2)=0.08; P(S1,S2,S3|D)=0.000032; how would you use this information to make a “better” prediction of P(D|S1,S2,S3)? [3]

c) How can the discrepancy in the prediction in the cases a) and b) be explained? Why are the numbers you obtain different? What does this discrepancy tell you about naïve Bayesian systems in general?

5) Computations in Belief Networks (w=5)

Assume that the following Belief Network is given that consists of nodes A, B, C, D, and E that can take values of true and false.

a) Using the given probabilities of the probability tables of the above belief network (D|A,B; E|C,D; A; B; C;…) give a formula to compute P(E|B). Explain all nontrivial steps you used to obtain the formula! [9]

b) Is A**|**D,E d-separable from C**|**D,E? Give reasons for your answer! [2]

c) Is A,B|∅ d-separable from C|∅? Give reasons for your answer! ∅:=”no evidence” [2]

6. Belief Network Design & Using a Belief Network Tool (w=9)



Assume we have 3 astronomers in different parts of the world who make measurements M1, M2, and M3 of the number[[3]](#footnote-3) of stars N in some region of the sky. Normally, there is a probability of 0.05 that the astronomer counts a single star twice (overcounts by one star; you can assume that the three astronomers never undercount; moreover, if there is no star visible (N=0) the astronomer never overcounts). Moreover, there is a 10% probability (P(Fi=1)=0.1 for i=1,2,3) that a telescope is out of focus (represented using random variables F1, F2, and F3), in which the astronomer undercounts by 2 or more stars (e.g. if N is 4 and her telescope is out of focus, the astronomer will count 2, 1 or 0 stars; you can assume if information is missing that each case has the same probability). Design a belief network, and compute the probability of the other variables assuming the following pieces of evidence are given (feel free to use *Netica (*<http://www.norsys.com/download.html> ) or any another belief network tool to compute your answer!):

1. M1=3 M2=3 M3=1
2. M1=4 M2=4 M3=1
3. M1=6 M2=6 M3=4
4. N=4, M2=1, M3=0
5. M1=0 M2=3 M3=4
6. M1=3 M2=4 M3=4
7. N=3 F1=0 F2=0 F3=1
8. M1=4 M2=5 F3=1

Submit the complete Belief Network you created, and the findings you obtained for the eight cases listed above!

1. No collaboration with your classmates are allowed for solving problems 1-5. [↑](#footnote-ref-1)
2. You are allowed to solve this problem with your teammates in Project2; please submit only one solution per group for problem 6! [↑](#footnote-ref-2)
3. You can assume that N is limited to 5—but the astronomer do not know that: M1,M2,M3,M4 are therefore limited to 0 through 6. [↑](#footnote-ref-3)