

6.034 QUIZ 2 SOLUTIONS
FALL 2000

(Preliminary – without Explanations)

Problem1: Miscellaneous(30p oints)

This problem is first caused by the problem where it is judged that the correct answer is four times as long. The correct answer is **inglep** because it has the most complete following fragments. All multiple choice answers are rejected.

Progressive opening, also known as iterative opening, works either for or against the cause:

- Alpha-beta allows you to get a wide range of answers.
- The anchoring factor varies from left to right.
- Almost none of the answers are given in the final answer.
- **Almost all answers are given in the final answer.**
- All of the above.
- None of the above.

Alpha-beta:

- Double the number of iterations.
- Is lower than the number of iterations.
- Is incompatible with the number of iterations.
- Is incompatible with the iterative opening.
- All of the above.
- **None of the above.**

The topological sorting algorithm was developed to:

- Improve the number of iterations.
- Ensure the order of termination is the same as the join principle.
- Deal with loops in the inheritance tree.
- **Honor the principles.**
- All of the above.
- None of the above.

Frames are used to:

- Enable the understanding of the situation.
- Enable the understanding of the situation.
- Enable the understanding of the situation.
- Enable the understanding of the situation via inheritance.
- **All of the above.**
- None of the above.

Ake yvi rtueof t het ransition-spacer epresentationi st hati t:

- Subsumest hematicr olef ramesa ndpr imitive-actf rames.
- Canbe t ranslatedt or elational-databaser ecords.
- Enablest hede scriptionof s tates.
- Facilitatest heund erstandingof m etaphors.
- Allof t hea bove.
- Noneof t hea bove.

Ake yvi rtueof t het hematic-rolf rame representationi st hati t:

- Expressesa lla ctionsi nt erms of a f ewpr imitivea cts.
- Enablesde scriptiona tt hes toryl evel.
- Focusesonv ariable-valuec hanges.
- Capturesc ausalr elations.
- Allof t hea bove.
- Noneof t hea bove.

Ake yvi rtueof s emantic-transition-tree grammars ist hatt hey:

- Reducet henum berof w ordst hatne edt obe und erstood.
- Simplify grammarc onstructionb ys ubstitutingr ecursionf ore xplicitl oops.
- Exploitt her etea lgorithm.
- Exploitt ransition-spacer epresentation.
- Allof t hea bove.
- Noneof t hea bove.

Naturall angugeda tabasei nterfacesw orkbe causepr ocessednounphr asesm ostof ten become:

- Relationalj oinope rations.
- Relationalselecionope rations.
- Relationalpr ojectionope rations.
- Relationalsortingope rations.
- Allof t hea bove.
- Noneof t hea bove.

The purpose of crossover in genetic algorithms is to:

- Increase diversity.
- Model natural mutation.
- Change selection probabilities.
- Reduce the number of genotypes.
- All of the above.
- None of the above.

In neural networks:

- Biological neural networks are modeled accurately.
- Sigmoid thresholds were introduced to avoid overfitting.
- Overfitting occurs if there are too few training cycles.
- The computation required by backpropagation per training cycle is proportional to n^2 , where n is the number of nodes.
- All of the above.
- None of the above.

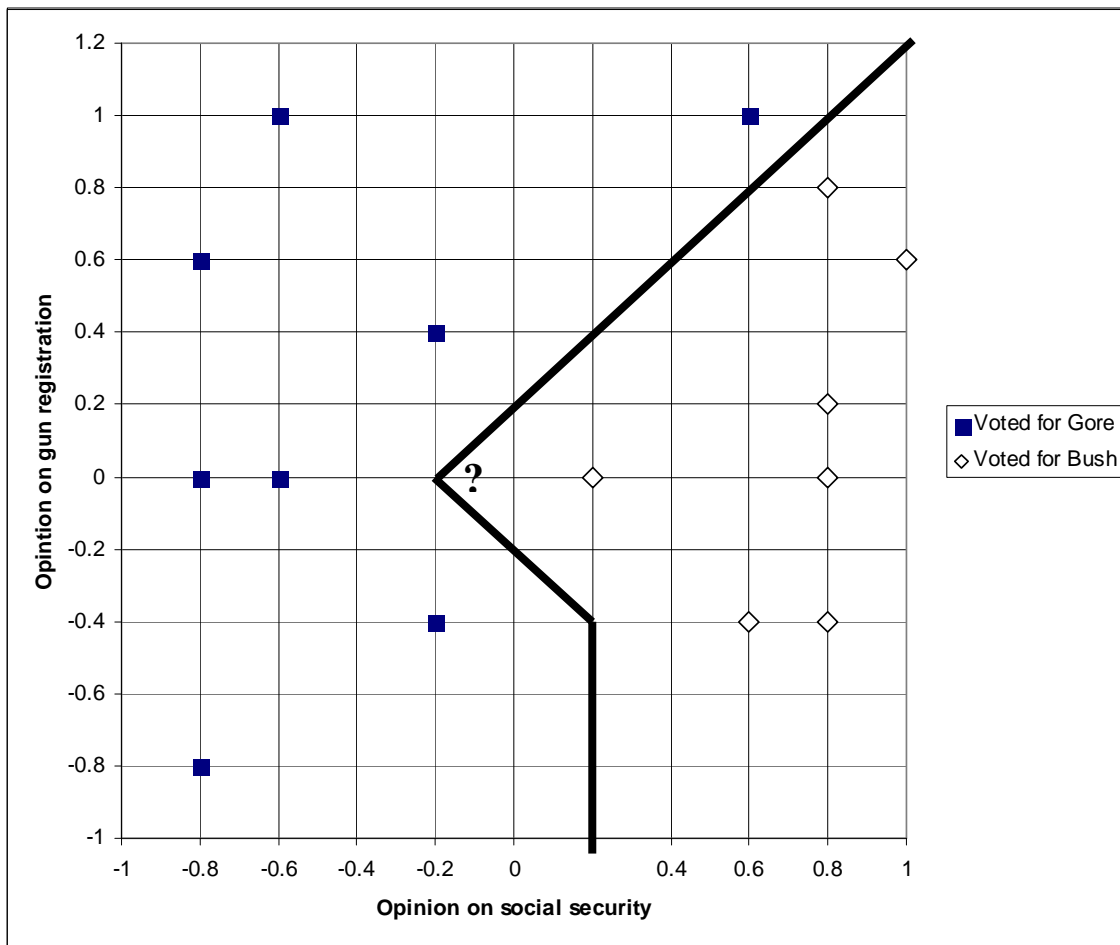
Problem 2: Nearest Neighbor, Decision Trees and Support Vectors (30 points)

Congress has decided to ask voters a few key questions so as to predict how each will vote. This will, of course, save everyone the troublesome and time-consuming practice of actually having to examine the ballots to figure out the election result.

They decide to start with just two questions:

- a) On a scale of -1 (strongly disagree) to $+1$ (strongly agree) how do you feel about privatizing social security?
- b) On a scale of -1 to $+1$, how do you feel about registering handguns?

The training sample is shown below, with 15 individuals plotted according to how they feel about these two issues, with a dark square (for Gore) and a light diamond (for Bush) indicating how they voted for president. There is also a question mark “?” on the plot, indicating one of the infamous undecided voters about which candidate has been aided during this election. We’ll call him Mr. Undecided.



Part A : Nearest Neighbor (10 points)

- 1) What would the nearest neighbor predict about the vote of Mr. U undecided, assuming he uses the standard Euclidean distance metric? (Your answers should be either Gore or Bush.)
- 2) On the plot above, carefully draw with pencil the nearest neighbor would indicate as separating the Gore part of the samples space from the Bush part.

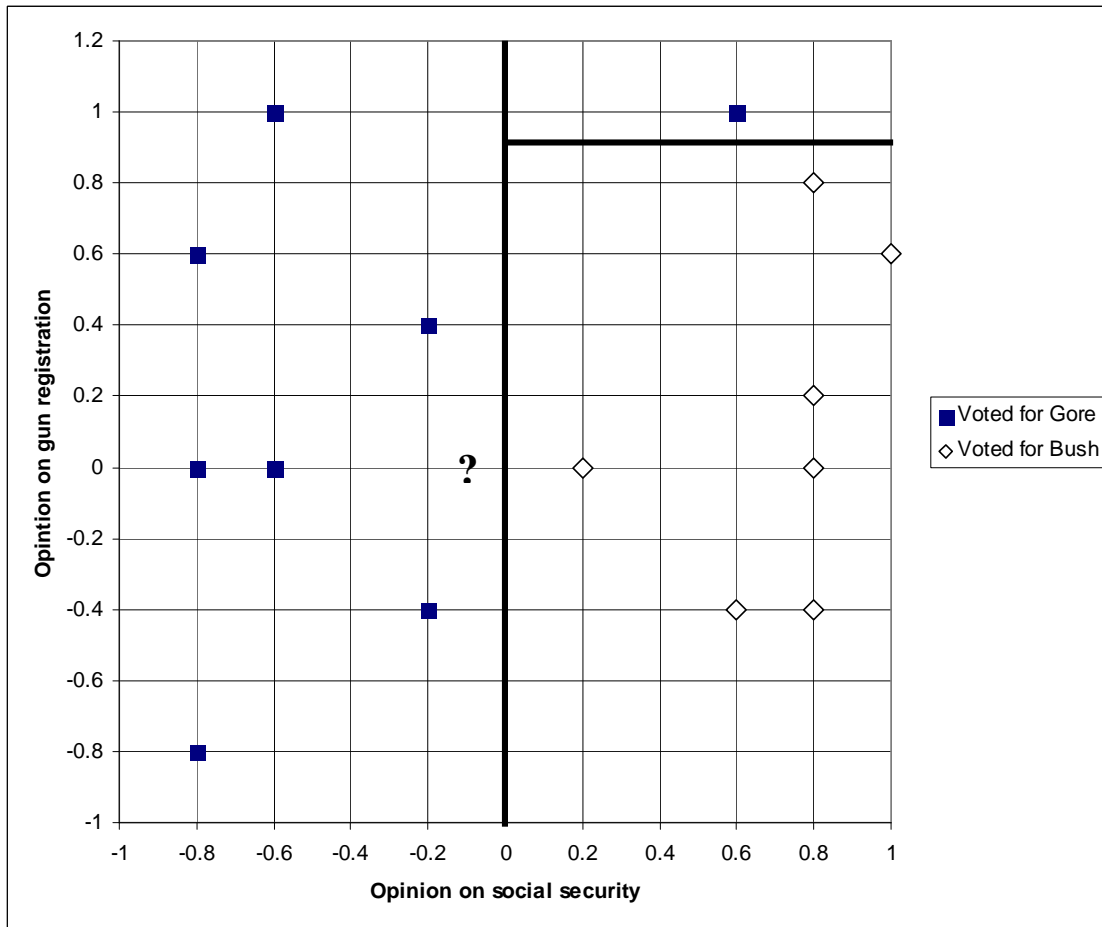
- 3) What would the nearest neighbor predict about the vote of Mr. U (using the same Euclidean metric)?

Gore

- 4) It turns out that here we have a question that was asked: "How do you feel about lowering the pay for Congressmen/women?" The question was not included in the public release database, (according to the politicians who controlled the release), that data will not be useful in making a decision. We have a question, you find out that the actual problem was that all the answers were strongly clustered near the end of the scale. You are brought in as a consultant and suggest that:
- a) The politicians are correct; that data will not be useful.
 - b) The data are still not useful, you just need to **divide** all the values by the mean.
 - c) The data are still not useful, you just need to **subtract** the mean from all the values, then **multiply** all the values by the standard deviation of the values.
 - d) The data are still not useful, you just need to **divide** by the standard deviation.
 - e) The data are still not useful, but none of the choices of the data above are correct.

B: Identification Trees (8p oints)

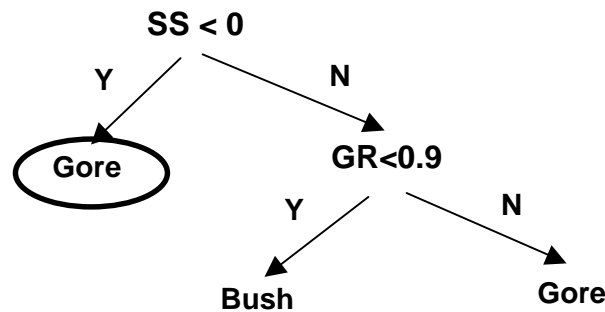
(Wer epeatt hes ameda tahe ref or yourc onvenience.)



Things seem to be going along well, when suddenly, Republican leaders appear on the scene and suggest that the environment is being wrecked by wasting precious time and pace. They suggest signing the ID Act instead.

- 1) You decide that your first step is opinion on social security < 0. But as you know, you need to determine the average order of the steps produced by this set of seven different things. What is the average order? (Your answer can include the logarithm operator, you need not simplify your expression.)

- 2) Youde cidet hati tl ooksg ood,s o youde cidet oc ompletet hede cisiont ree. D rawt he IDt ree ands pecifya llt ests.



- 3) Whatdoe s yourt reepr edicta bouthow M r.U ndecidedw illvot ef orpr esident?C ircle thec orrespondingnod eo n your IDt ree.

PartC : More VoterQ uestions(4p oints)

Mr.G ore,h avingi nventedt he Internet, claimst oknow a t hingo rt woa boutt echnology. Hes ayst hatw e're askingw ayt oof ewqu estionsof t hevot ers,a ndi ndicatest hatt og et a decentpr edictivea bility wes houllda skt hema tl east100que stions.Y ouc omeupw ith 100que stions,a ndw hiley ouw orkedha rda ti t, youdon' tt hinka ll100que stionsa re goingt o give youpr edictivei nformationa boutt hevot er.N evertheless, you forgea head andt ryus ingbot hne arestne ighbors andi dentificationt rees.Y ouri nitiale xperiments indicate(circlet hem ostl ikelyr esult):

- a) Botht echniquesw orkw ella ndw orka boutt quallyw ell.
- b) Nearestne ighborsw orks muchbe tttert hani dentificationt rees.
- c) Identificationt reesw ork muchbe tttert hanne arest neighbors.
- d) Neitherw orks well.

Part D: Support Vector Machines (8 points)

All of this is about the new hidden units in the neural network. The demonstration goes on out into the street, with the people carrying signs that say

Support
Vector Machines

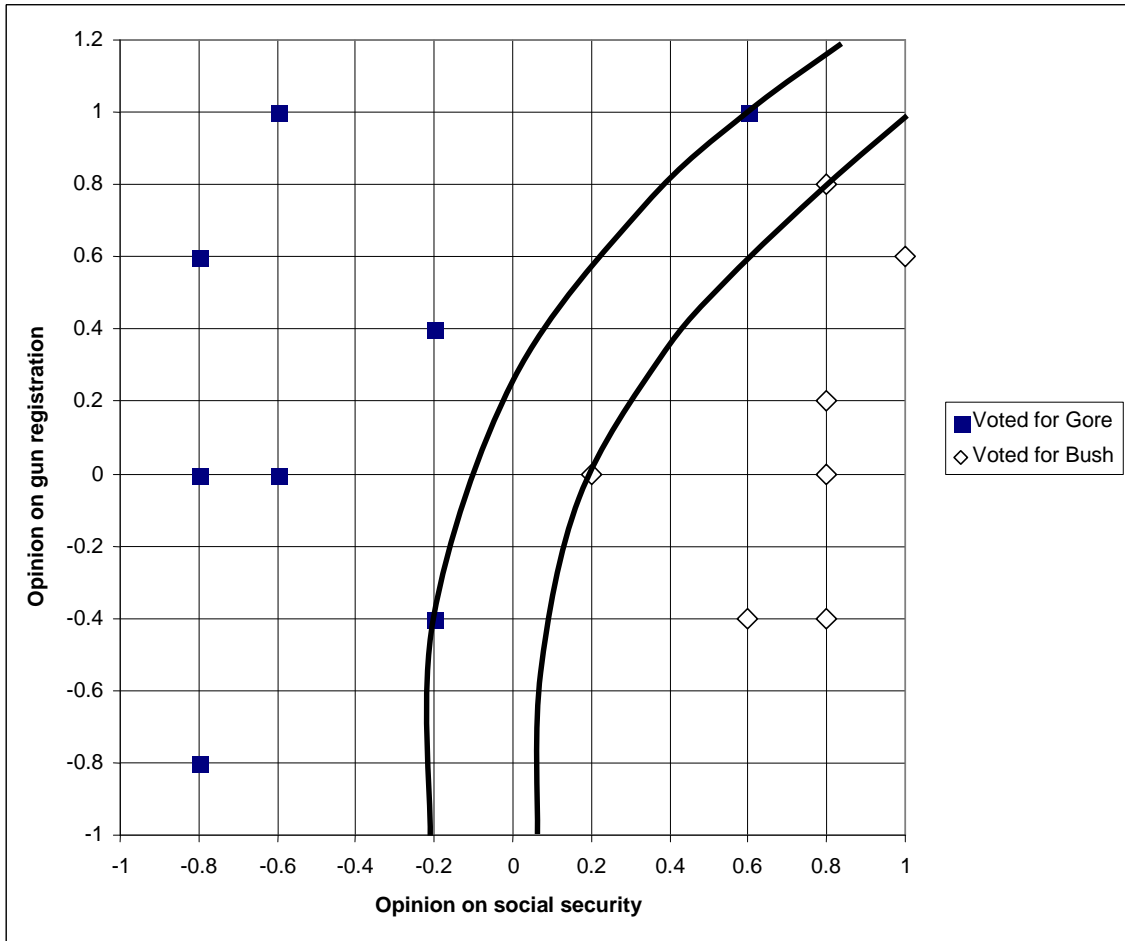
You're a bit confused and ask a colleague, "What's a support vector machine?" He explains that you've probably heard this term incorrectly, it's called a support vector machine. This brings you all back to you, and you indicate your understanding by answering the following question. (At the discretion of the grader, it is allowed to be low for your convenience.)

Your friend experiments with a radial-basis-function kernel,

$$K(\mathbf{x}_1, \mathbf{x}_2) = e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}}$$

Your experiments succeed, with a small number of samples merging into support vectors, and all samples correctly providing a value of either 0 or 1.

1. Sketcht hede cisionboun daryi nt hedi agrambe low.



2. Check all hes tatementst hata rec orrect:

Therei sa s igmav alues ucht hata llt heG ore/Bushs amplesw illbe come supportve ctors.

Therei sa s igmav alues ucht hatt heG ore/Bushs amplesw illnot be separable.

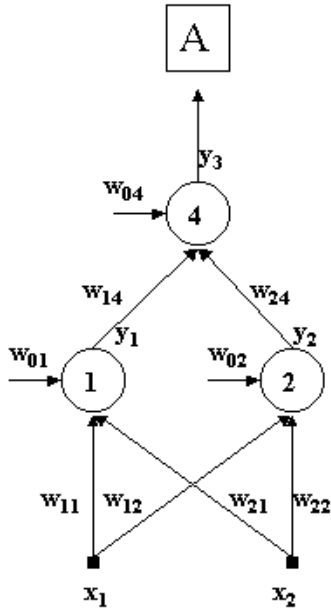
Therei sa s igmav alues ucht hatove rfittingw illoc cur.

Ift her adial-basis-functionke rneli sr eplacedb y a dotpr oductke rnel, $x_1 \cdot x_2$,t heG ore/Bushs amplesw illbe s eparable.

Problem3: Neural Networks (40 points)

PartA (24 points):

Consider the following neural network. All three of the neurons are perceptrons with a sigmoidal transfer function; that is, all three use the step function as their transfer function. Assume all inputs to the hidden layer have a bias input of 1. Implementing weights, such as w_{01} and w_{02} are -1.



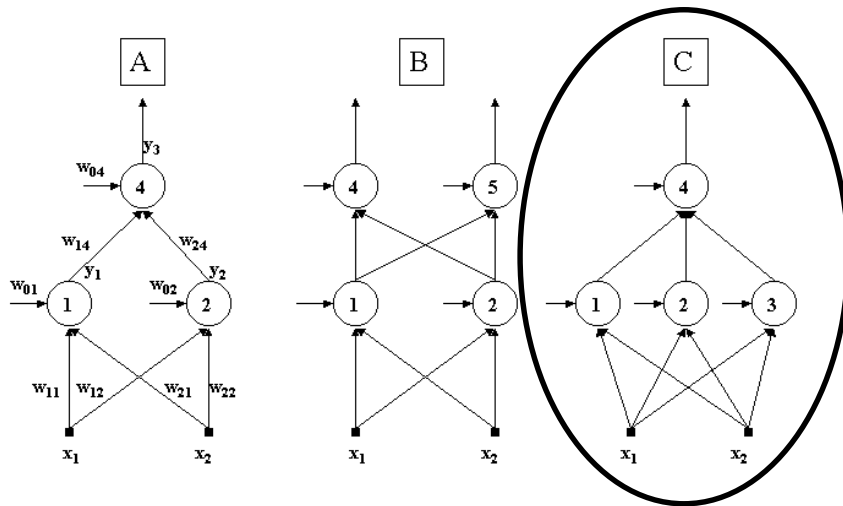
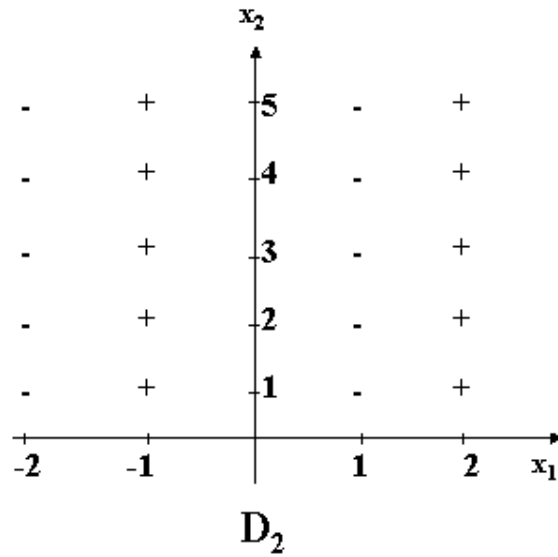
A.1 (10 points)

Give values for the missing weights of net A so that it correctly classifies the Data in D_1 below. The output for each instance should be 0 if or instances labeled - and 1 if or instances labeled +.

<p style="text-align: center;">D_1</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>$w_{01} = -1.5$</td></tr> <tr><td>$w_{11} = -1$</td></tr> <tr><td>$w_{21} = 0$</td></tr> <tr><td>$w_{02} = -1.5$</td></tr> <tr><td>$w_{12} = 1$</td></tr> <tr><td>$w_{22} = 0$</td></tr> <tr><td>$w_{04} = 1.5$</td></tr> <tr><td>$w_{14} = 1$</td></tr> <tr><td>$w_{24} = 1$</td></tr> </table>	$w_{01} = -1.5$	$w_{11} = -1$	$w_{21} = 0$	$w_{02} = -1.5$	$w_{12} = 1$	$w_{22} = 0$	$w_{04} = 1.5$	$w_{14} = 1$	$w_{24} = 1$
$w_{01} = -1.5$										
$w_{11} = -1$										
$w_{21} = 0$										
$w_{02} = -1.5$										
$w_{12} = 1$										
$w_{22} = 0$										
$w_{04} = 1.5$										
$w_{14} = 1$										
$w_{24} = 1$										

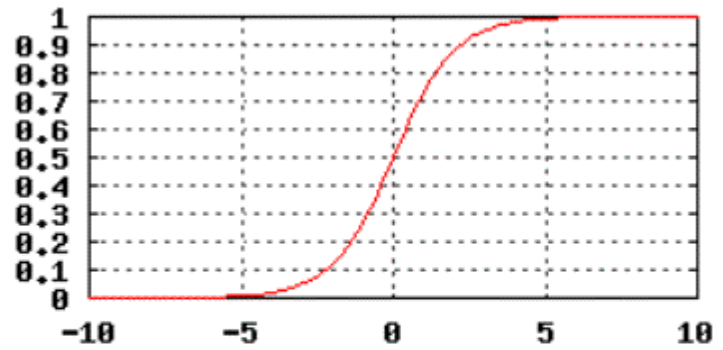
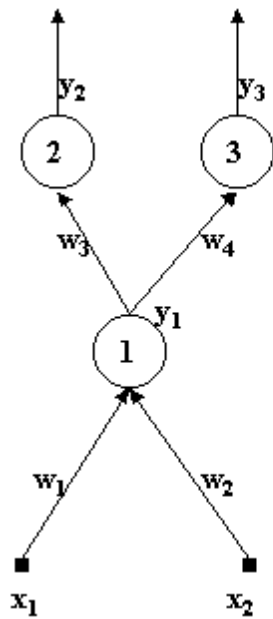
A.3(8p oints)

Consider the following data set, which is somewhat different from D_1 :



Circle the best of the three architectures, if any, that can correctly classify this data set. Assume that all architectures use sigmoidal perceptron units.

Part B (16p oints)



Sigmoid Values

S(-5)= 0	S(-4)=0.02	S(-3)=0.04	S(-2)=0.12	S(-1)=0.27	S(0)=0.5
S(5)=1	S(4)=0.98	S(3)=0.96	S(2)=0.88	S(1)=0.73	

B.1(6poi nts)

Assuming the activation function is the sigmoid function with a fixed threshold of 0.0, compute (approximately), given that the inputs are $x_1=2, x_2=-2, w_1=2, w_2=-2, w_3=4, w_4=0$.

- $y_1 = 1/2$
- $y_2 = 0.88$
- $y_3 = 1/2$

B.2(8p oints)

You are confronted with two situations, and you are to determine whether the weights will go up or down. In both situations, you are to use the same two weights from B.1, with the usual sigmoidal neurons.

In **situation1**, you are to assume that the desired outputs for the neurons

$$x_1 = 2,$$
$$x_2 = 2$$

are

$$y_2 = 0,$$
$$y_3 = 1,$$

that the learning rates are small, and that the weights are initially 1.

In **situation2**, you are to assume that the desired outputs for the neurons are

$$y_2 = 1,$$
$$y_3 = 0,$$

Fill in the cells of the table below with **up** or **down**, as appropriate.

	Situation1	Situation2
w₁	↓	↑
w₂	↓	↑
w₃	↓	↑
w₄	↑	↓