CIS 472/572 Midterm Exam

Mamtaj Akter

TOTAL POINTS

47 / 48

QUESTION 1

PROBLEM 1: PERCEPTRONS 6 pts

1.1 The convergence rate 2 / 2

✓ **- 0 pts Correct**
- **- 0.5 pts** Vague answer
- **- 0.5 pts** Confusing answer
- **- 1 pts** Wrong reason
- **- 2 pts** Incorrect
- **- 2 pts** Blank
- **- 0 pts** Skipped

1.2 The final perceptron accuracy on the training data could change 1.5 / 2

- **- 0 pts** Correct

✓ **- 0.5 pts Vague or confusing**
- **- 1 pts** Wrong reason
- **- 2 pts** Incorrect
- **- 2 pts** Blank
- **- 0 pts** Skipped

1.3 The final perceptron accuracy on a separate, unseen test set 1.5 / 2

- **- 0 pts** Correct

✓ **- 0.5 pts Explanation is vague, confusing, or incomplete.**
- **- 1 pts** Wrong reason
- **- 2 pts** Incorrect
- **- 2 pts** Blank
- **- 0 pts** Skipped

QUESTION 2

2 PROBLEM 2: PERCEPTRON UPDATES 6 / 6

✓ **- 0 pts Correct**
- **- 1 pts** Minor error
- **- 2 pts** Medium error
- **- 4 pts** Major error
- **- 6 pts** Incorrect
- **- 6 pts** Blank
- **- 0 pts** Skipped

QUESTION 3

PROBLEM 3: REPRESENTATION 6 pts

3.1 Draw a decision tree 3 / 3

✓ **- 0 pts Correct**
- **- 1 pts** Minor error
- **- 2 pts** Medium error
- **- 3 pts** Incorrect
- **- 3 pts** Blank
- **- 0 pts** Skipped

3.2 Specify parameters for a linear classifier... 3 / 3

✓ **- 0 pts Correct**
- **- 0.5 pts** Minor error
- **- 1.5 pts** Medium error
- **- 3 pts** Incorrect
- **- 3 pts** Blank
- **- 0 pts** Skipped

QUESTION 4

PROBLEM 4: LEARNING POWER 6 pts

4.1 3-nearest neighbor 2 / 2

✓ **- 0 pts Correct**
- **- 1 pts** Incorrect or inadequate explanation.
- **- 2 pts** Incorrect
- **- 2 pts** Blank
- **- 0 pts** Skipped

4.2 Logistic regression: 2 / 2

✓ - **0 pts** Correct

- **1 pts** Wrong explanation
- **2 pts** Incorrect
- **0 pts** Skipped

4.3 Neural network with a single hidden layer, 2 / 2

✓ - **0 pts** Correct

- **1 pts** Wrong explanation
- **2 pts** Incorrect
- **0 pts** Skipped

QUESTION 5

PROBLEM 5: LINEAR SUPPORT VECTOR MACHINES 6 pts

5.1 Draw the decision boundary... 3 / 3

✓ - **0 pts** Correct

- **1 pts** No support vectors identified
- **3 pts** Incorrect
- **3 pts** Blank
- **0 pts** Skipped

5.2 Calculate the leave one out cross validation error... 3 / 3

✓ - **0 pts** Correct

- **1 pts** Minor error
- **2 pts** Only one of the support vectors is mispredicted when left out.
- **2 pts** Not all examples will be predicted correctly.
- **3 pts** Incorrect or blank
- **0 pts** Skipped

QUESTION 6

PROBLEM 6: KERNELS 6 pts

6.1 Linear kernel, 2 / 2

✓ - **0 pts** Correct

- **1 pts** Partly wrong
- **2 pts** Incorrect or blank
- **0 pts** Skipped

6.2 Quadratic kernel 2 / 2

✓ - **0 pts** Correct

- **1 pts** Partly wrong
- **2 pts** Incorrect or blank
- **0 pts** Skipped

6.3 Cubic kernel 2 / 2

✓ - **0 pts** Correct

- **1 pts** Partly wrong
- **2 pts** Incorrect or blank
- **0 pts** Skipped

QUESTION 7

7 PROBLEM 7: GRADIENT DESCENT 6 / 6

✓ - **0 pts** Correct

- **1 pts** Minor error
- **2 pts** Medium error
- **4 pts** Large error
- **6 pts** Incorrect or blank
- **0 pts** Skipped

QUESTION 8

PROBLEM 8: APPLICATIONS 6 pts

8.1 Give one reason why a neural network is a good choice for this problem. 2 / 2

✓ - **0 pts** Correct

- **1 pts** Partially correct
- **2 pts** Incorrect or blank
- **0 pts** Skipped

8.2 Give one reason why a decision tree might be a bad choice for this problem. 2 / 2

✓ - **0 pts** Correct

- **1 pts** Partially correct
- **2 pts** Incorrect or blank
- **0 pts** Skipped

8.3 Give one reason why logistic regression

might be a bad choice for this problem **2 / 2**

- ✓ **- 0 pts** **Correct**
- **- 1 pts** Partially correct
- **- 2 pts** Incorrect or blank
- **- 0 pts** Skipped

EXAM: WINTER 2017
CIS 472/572
INSTRUCTOR: DANIEL LOWD

March 1, 2017

The exam is closed book and open notes (1 page, handwritten except with prior permission). Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, continue on the back of the page. Undergraduates only: You may skip one set of questions (either Problem 1, 2, 3, 4, 5, 6, 7, or 8). Please write down on the front of your test which problem you are choosing to skip. You will receive full credit on the skipped question set.

NAME Mamtaj Akter

PROBLEM 1: PERCEPTRONS

Suppose you learn a perceptron on a linearly separable dataset by running the perceptron learning algorithm until convergence. Which of the following could be affected by changing the order of the training examples? For full points, **you must briefly explain each answer** (one sentence is enough).

1. (2 points) The convergence rate (number of iterations before convergence) could change. True or False. **Explain.**

   True. Because randomly picked examples leds the model to converge faster.

2. (2 points) The final perceptron accuracy on the training data could change. True or False. **Explain.**

   False, perceptron accuracy won't change as the training data is same.

3. (2 points) The final perceptron accuracy on a separate, unseen test set could change. True or False. **Explain.**

   True, perceptron doesn't guaranty same accuracy for training and test dataset.

PROBLEM 2: PERCEPTRON UPDATES

You are training a classifier to distinguish between bubonic plague (+1) and the flu (-1).

Patient	chills X_1	fever X_2	cramps X_3	seizures X_4	gangrene X_5	Disease Y
1.	1	1	0	1	1	1
2.	1	1	0	0	0	-1
3.	1	1	1	1	0	1
4.	1	0	1	0	0	-1

1. (6 points) Show the weights and bias (w and b) obtained by running perceptron algorithm on this dataset for one iteration. (Here, "one iteration" means going over all of the examples once, in the order shown above.)

$\rightarrow w = \langle 0,0,0,0,0 \rangle$, $b=0$

$y(wx+b) = 0 \not> 0$ so update $w = w + xy$, $b = b + y$

$\rightarrow w = \langle 1, 1, 0, 1, 1 \rangle$ $b = 1$

$y(wx+b) = (-1)(1+1+0+0+0+1) = -3 \not> 0$ so upc

$\rightarrow w = \langle 0, 0, 0, 1, 1 \rangle$, $b = 0$

$y(wx+b) = 1(0+0+0+1+0+0) = 1 > 0$ so no up

$\rightarrow w = \langle 0, 0, 0, 1, 1 \rangle$, $b=0$

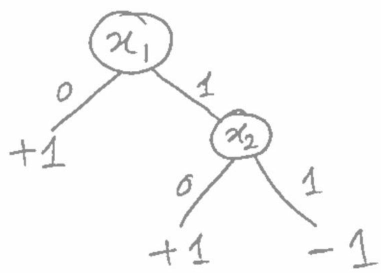
$y(wx+b) = (-1)(0+0+0+0+0+0+0) = 0 \not> 0$ so upda

$w = \langle -1, 0, -1, 1, 1 \rangle$, $b = -1$

PROBLEM 3: REPRESENTATION

Let x_1 and x_2 be two binary-valued attributes, which can take on values of 0 or 1. Consider the NAND function, $y = \neg(x_1 \wedge x_2)$, which is -1 if both $x_1 = 1$ and $x_2 = 1$ and +1 otherwise.

1. (3 points) Draw a decision tree that represents this function.

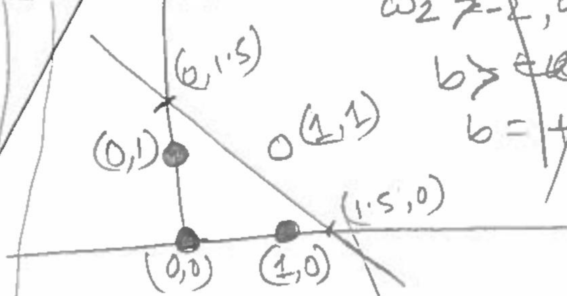


x_1	x_2	$x_1 \wedge x_2$	$\neg(x_1 \wedge x_2)$
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	-1

2. (3 points) Specify parameters for a linear classifier that represents this function. (You do not need to find a maximum margin separator – any separator will do. Do not create any additional attributes.) $\boxed{w_1 = -1,\ w_2 = -1,\ b = 1.5}$ $\rightarrow$ $w_1 > -2$, $w_1 =$
$w_2 > -2$, $w_2 =$
$b >$
$b = 1$

$wx + b = +1$ in $(0, 1)$, or, $(1, 0)$
$\Rightarrow w \cdot 1 + b = +1$
$\Rightarrow b + w = 1$
$wx + b = -1$ in $(1, 1)$
$\Rightarrow w \cdot 1 + b = -1$
$\Rightarrow b + w = -1$

$w \cdot 0 + 1.5 > 1$
$(-1) \cdot 1 + 1.5 > 1$



$0 + 0 + 1.5 > 0$
$0 - 1 + 1.5 > 0$
$-1 + 0 + 1.5 > 0$
$-1 - 1 + 1.5 < 0$

$1 \cdot (-1) + 1.5$

$w = -1$,
$b = 1.5$

$(-1) \cdot 0 + 1.5 > 1$
$(-1) \cdot 1 + 1.5 > 1$
$(-1) 0 + 1.5 > 1$

$\frac{y - 1.5}{1.5 - 0} = \frac{x - 0}{0 - 1.5}$
$\Rightarrow \frac{y - 1.5}{1.5} = \frac{x}{-1.5}$
$\Rightarrow -1.5y + 2.25 = 1.5x$
$\Rightarrow -1.5y = 1.5x - 2.25$
$\Rightarrow -y = x - 1.5$
$\Rightarrow y = -x + 1.5$

PROBLEM 4: LEARNING POWER

Which of the following classifiers will have zero training error (under 0/1 loss) on the following dataset? Circle YES for classifiers that will have zero error and NO for those that will not. Give a short explanation (one sentence) for each answer.

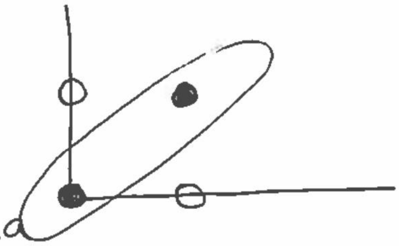
X1	X2	Category	
1	1	0	2 1s ~~1~~ 0
1	0	1	2 0s ~~0~~
0	1	1	2 0s
0	0	0	2 1s

1. (2 points) 3-nearest neighbor: YES or NO? Explain.

NO. because ~~all~~ each of the ~~dat~~ exampl will always get ~~opposite cla~~ two opposite class. So, each example will get always wrong neighbor.

2. (2 points) Logistic regression: YES or NO? Explain.

No. — Not linearly classified. need to draw a parabola



3. (2 points) Neural network with a single hidden layer, and only one node in the hidden layer: YES or NO? Explain.

No, It's not possible to represent such complex boolean function (not linearly separable) with a single hidden layer which has only one hidden node.

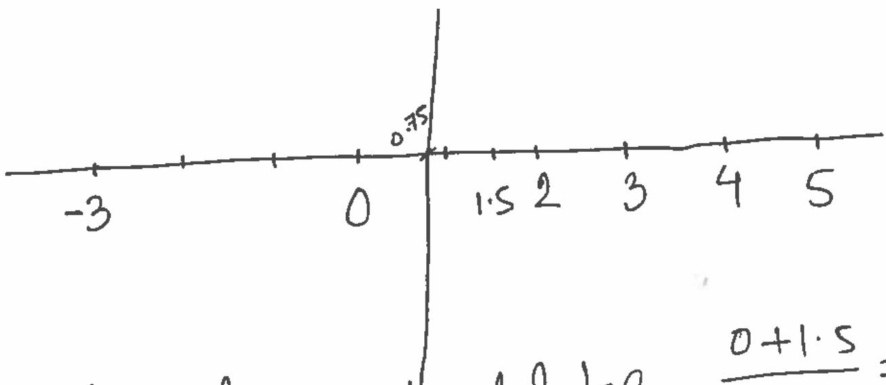
PROBLEM 5: LINEAR SUPPORT VECTOR MACHINES

Consider the following 1-dimensional data:

	−	+	+	+	+	+	+
x	-3	0	1.5	2	3	4	5
Class	−	−	+	+	+	+	+

1. (3 points) Draw the decision boundary of a (hard-margin) linear support vector machine on this data and identify the support vectors. (The boundary should be a single point on the number line, separating the positive and negative classes.)

support vectors are 0, 1·5



decision boundary should be $\frac{0+1\cdot5}{2} = 0\cdot75$ p

2. (3 points) Calculate the leave one out cross validation error for this SVM on the data set. (That is, how many individual points would be predicted incorrectly if they were removed from the training data?)

LOOCV

only 1 error, when 0 is removed, it gets 1·5 as the closest neighbor, which has +ve class, so wrong classification.

PROBLEM 6: KERNELS

Consider a kernelized SVM with $b = 0$ and the following instances and weights:

instance ($x^{(i)}$)	label ($y^{(i)}$)	weight (α_i)
(0,0,0)	+1	1.0
(1,1,1)	-1	1.0

What is the predicted label for the instance $x = (-1, -1, -1)$ under each kernel? (NOTE: Be careful with positive and negative signs! I recommend showing your work in order to have a chance at partial credit.)

1. (2 points) Linear kernel, $K(x, x') = x \cdot x'$. $\sum y \alpha (x \cdot x') + b$

$1 \cdot 1 \cdot (0+0+0) + 1 \cdot (-1)(-1-1-1) + 0 = 3 + 0 > 0$

so, predicted lablel will be +1.

2. (2 points) Quadratic kernel, $K(x, x') = (1 + x \cdot x')^2$ $\sum y \alpha (1 + x \cdot x')^2 + b$

$1 \cdot 1 \cdot (1+0+0+0)^2 + 1 \cdot (-1) \cdot (\cancel{1} + (\cancel{-1}) - 1 - 1)^2 + 0$

$\cancel{= 1 + 8 + 0 = 9 > 0}$ $= 1 + (-4) + 0 = -3 \not> 0$

so, predicted label will be ~~+1~~. −1

3. (2 points) Cubic kernel, $K(x, x') = (1 + x \cdot x')^3$ $\sum y \alpha (1 + x \cdot x')^3 + b$

$1 \cdot 1 \cdot (1+0+0+0)^3 + 1 \cdot (-1) \cdot (\cancel{1} + (\cancel{-1}) - 1 - 1)^3 + 0$

$= 1 + (-1)(-8) = 9 > 0$

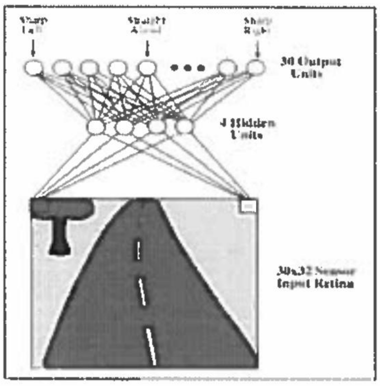
so predicted label will be +1.

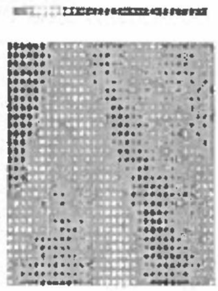
PROBLEM 7: GRADIENT DESCENT

1. (6 points) Starting at each location marked with a star, draw the path that gradient ascent (that is, following the gradient uphill) would take until it converges to a local optimum. Assume that the step size is small enough for a smooth path. Your path does not need to be perfect.

PROBLEM 8: APPLICATIONS

In 1989, 20 years before Google's self-driving car project, researchers at Carnegie Mellon developed an autonomous driving system called ALVINN. ALVINN uses a simple neural network to pick the steering direction based on real-time video input. The video input is a 30x32 pixel camera image and the output is the steering direction. ALVINN was trained using simulated road images along with the recommended steering direction.





1. (2 points) Give one reason why a neural network is a good choice for this problem.

Neural network is a good choice for this particular problem as it is good at handling non-linear complex functions and image processing/visualization which could be very nonlinear. And its also efficient at predictions.

2. (2 points) Give one reason why a decision tree might be a bad choice for this problem.

driving decisions relies on almost all pixels here which requires a huge (sparse) tree to test. So, its a bad idea to use decision tree here.

3. (2 points) Give one reason why logistic regression might be a bad choice for this problem.

Logistic regression is linear, so ~~as~~ each pixels are same level of importants, using a linear model is not enough to serve the purpose here.