CIS 391
Introduction to Artificial Intelligence
Practice Midterm II
With Solutions

(From old CIS 521 exam)

Problem	Points	Possible
1. Perceptrons and SVMs		20
2. Probabilistic Models		20
3. Markov Models		20
4 Propositional Logic		20
5. Deductive Inference		20
TOTAL		**100**

1 Perceptrons and Support Vector Machines [20 points]

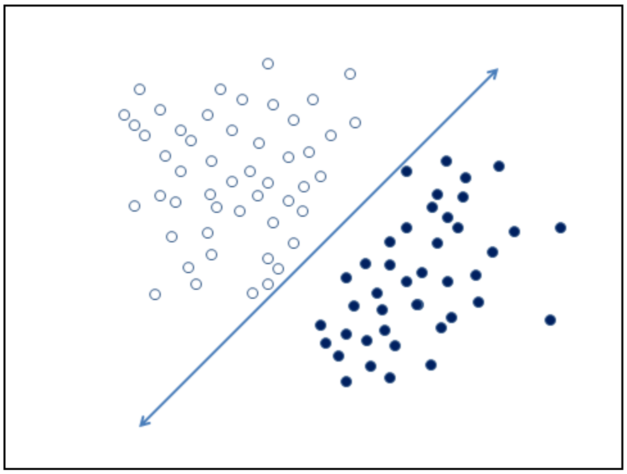
The pictures below represent simple 2-dimensional classification problems, where the dots in each picture represent the set of training examples, the kind of dot represents the class, and the arrow represents a line which may or may not separate the two kinds of dots. Somewhat more formally, the set of dots in each represents a 2 dimensional set $\{x_i, y_i\}$ of training examples, where each x_i is a 2 dimensional vector representing the location of the point, and y_i is -1 for white dots and +1 for black dots (Note that mathematically, we would represent each line by a vector normal to it.)

1. [4 points] Could the picture immediately below represent a classifying hyperplane that results from training a perceptron on the training set shown ? Briefly explain your answer.

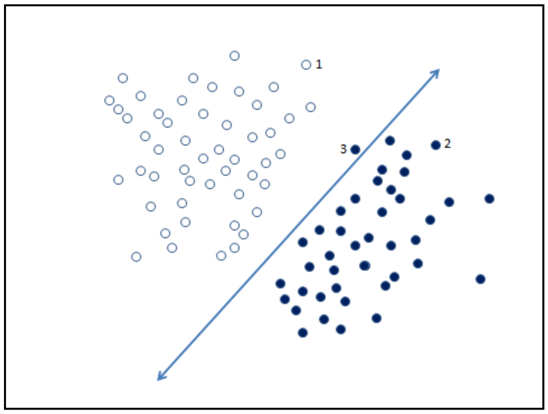
★ **SOLUTION:** Yes. This line is a hyperplane that separates the two classes of dots.

2. [4 points] Could the same picture represent a classifying hyperplane that results from training a Support Vector Machine on the training set shown? Briefly explain your answer.

★ **SOLUTION:** No, an SVM finds a hyperplane that maximizes the margins between the support vectors, i.e. it finds a separating hyperplane that is equidistant from the nearest points in both categories.



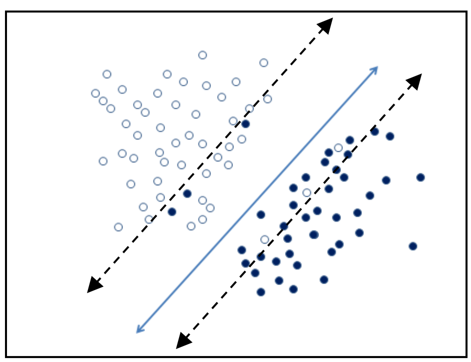
3. The picture immediately below represents a snapshot of a perceptron being trained on the data set shown, where the line shown is the algorithm's current attempt to find a separating hyperplane (here a line). Assume that the next three points to be considered by the perceptron algorithm in this iteration of the algorithm are, in order, the points labeled below as 1,2, and 3.



[4 points] Briefly describe how the perceptron algorithm adjusts the current hyperplane upon encountering each of 1,2 and 3.

★ **SOLUTION:** Since points 1 and 2 are both correctly classified by the current hyperplane, the hyperplane is left unaffected. Since point 3 is misclassified by this hyperplane , the perceptron rule rotates the hyperplane towards point 3.

4. Again, the picture immediately below represents a snapshot of a perceptron being trained on the data set shown, where the line shown is the algorithm's current attempt to find a separating hyperplane (here a line).



(a) [4 points] Describe the behavior of the perceptron algorithm as it runs over this data set.

★ **SOLUTION:** Since this data set is not linearly separable, the perceptron algorithm will be unstable, continually rotating the hyperplane first in one direction and then the other as it continually encounters misclassified points.

(b) [4 points] If we run the training algorithm on this data set k times, what can we say about the final classifying line the algorithm returns?

★ **SOLUTION:** We can't say very much. Depending on the order it encounters training vectors, it can end up anywhere between the two dashed lines above, roughly, i.e. it won't shift further than correctly classifying all of the white or all of the black dots, but could be anywhere in the middle.

2 Probabilistic Models [20 points]

1. Consider the following probabilistic universe associated with two Boolean random variables, *IsStudent* and *WearingBackpack* (whether or not a person is a student, and whether or not they are wearing a backpack):

	IsStudent	¬*IsStudent*
WearingBackpack	3/8	1/8
¬*WearingBackpack*	1/4	1/4

Each cell of this table represents the joint probability of the corresponding assignment to *IsStudent* and *WearingBackpack*.

(a) [3 points] What is the marginal probability that a person is a student?

★ **SOLUTION:** $P(S) = P(S|B) + P(S|\neg B) = 3/8 + 1/4 = 5/8$

(b) [3 points] Suppose I bet you with 1 to 2 odds that a given person is not a student (i.e. if the person is a student, I pay you $1, if not, you pay me $2). What is your expected net gain if you accept this bet? (If the value is positive, it is a good bet for you, if negative it is a bad bet for you.)

★ **SOLUTION:** Outcome is $1 \times (5/8) - 2 \times (3/8) = -1/8 \implies$ bad for you

(c) [6 points] Write down the conditional probability table (CPT) for the variable *IsStudent* given the value of *WearingBackpack*. Based on the CPT, are these variables independent?

★ **SOLUTION:** $P(S|B) = 3/4$, $P(S|\neg B) = 1/2$; these are not the same, so the variables are NOT independent.

COMMON MISTAKE: Just copying the *IsStudent* column of the table given with the problem is insufficient. You need to use the formula for conditional probability, like this:

$$P(S|B) = P(S \wedge B)/(P(B) = (3/8)/(3/8+1/8) = 3/4$$

2. Suppose we have a dataset with 3 binary features, X_1, X_2, and X_3, and we use Naive Bayes to classify a binary output Y with this data.

(a) [4 points] Write down the form of the joint probability model $P(X_1, X_2, X_3, Y)$ for this data using the Naive Bayes assumption.

★ **SOLUTION:** Naive Bayes assumes each feature is independent of the other features given the output Y. Thus, the joint probability is $P(Y)P(X_1|Y)P(X_2|Y)P(X_3|Y)$.

COMMON MISTAKE: Many people tried to write a formula for $P(Y|X_1, X_2, X_3)$ instead of $P(X_1, X_2, X_3, Y)$. These are not the same quantities. The former is the quantity you want to compute for predicting the value of Y, but the later (the joint probability) is the description of the full model.

(b) [4 points] How many parameters will we need to estimate to use Naive Bayes on this data? Hint: Don't count parameters whose value is implied by the fact that probability distributions must sum to 1.

★ **SOLUTION:** 1 for prior, plus 3 for each class = 7 total. That is, it is sufficient to estimate $P(Y = T)$, $P(X_i = T|Y = T)$, $P(X_i = T|Y = F)$, since all other values will be implied by the fact that probability distributions must sum to 1. No partial credit was given for this question.

3. Markov Models [20 Points]

A. [2 point] **TRUE or FALSE**: If we assume that a first order HMM describes a given sequence of n words and their tags, then we can always decompose the joint probability as

$$p(\mathbf{W},\mathbf{T}) = p(t_1)p(w_1|t_1)p(t_2|t_1)p(w_2|t_2) \ldots p(t_n|t_{n-1})p(w_n|t_n)$$

★ **SOLUTION:** True

Given the Markov model that you estimate from the two sequences below:

N	V	N		N	V	V	N
Markov	enjoyed	math		Mathematicians	enjoyed	growing	beards

B. [6 points] What is the probability $P(N|V)$?

★ **SOLUTION**: 2/3

C. [6 points] What is the probability $P(V|V)$?

★ **SOLUTION:** 1/3

D. [6 points] What is the probability of the tag sequence *"N V V V N"* being generated?

★ **SOLUTION:** $\pi(N) * P(V|N) * P(V|V) * P(V|V) * P(N|V)$ = 1 * 1 * 1/3 * 1/3 * 2/3 = 2/27

Problem: If students compute P(V|N) as ½, this counts as correct. Why? Count(NV)/Count(N)=2/4=1/2 and I didn't explain that the FINAL elements of the string shouldn't be counted in such a case. So 1/27 also gets full marks.

4. Propositional Logic [20 Points]

True/False and Multiple Choice (2 points each)

1. Consider a logic with only four propositional variables, *A, B, C* and *D*. How many models for this logic satisfy the sentence $(A \wedge B) \vee (B \wedge C)$?

(a) 3
(b) 4
(c) 8
(d) 16

★ **SOLUTION: BUG!!** There was a major typo here – the problem should have read $(A \wedge B) \vee (C \wedge D)$. Because of the bug, none of the solutions is correct, so full credit for any answer…..

2. The sentence $Smoke \Rightarrow Smoke$ is:

(a) Valid
(b) Unsatifiable
(c) Neither

★ **SOLUTION:** a

3. **True / False:** $(A \wedge B) \models (A \Leftrightarrow B)$

★ **SOLUTION: True**. The sentence $(A \Leftrightarrow B)$ will be true in any model in which $A \wedge B)$ is true.

4. **True / False:** $(A \Leftrightarrow B) \models (A \wedge B)$

★ **SOLUTION: False**, The sentence $(A \wedge B)$ will not be true in at least one model in which $(A \Leftrightarrow B)$ is true, namely any models in which A is False and B is False.

5. **True / False:** For any propositional sentences α, β, γ, if $(\alpha \wedge \beta) \models \gamma$ then either $\alpha \models \gamma$ or $\beta \models \gamma$ or both.

★ **SOLUTION: False.** Let $\gamma = (A \Leftrightarrow B)$. Then $(\alpha \wedge \beta) \models \gamma$ but it isn't true that $\alpha \models \gamma$ or $\beta \models \gamma$.

6. **True / False:** The resolution algorithm is both sound and complete for propositional logic.

★ **SOLUTION: True**

7. **True / False:** WalkSAT is both sound and complete for propositional logic.

★ **SOLUTION:** **False,** it's sound but incomplete

8. **True / False:** Any first order logic KB can be converted into a (potentially larger) equivalent propositional logic KB.

★ **SOLUTION:** Sigh… **True** if the world is finite, but **False** if it's infinite.

9. [4 points] Define **soundness** for an inference algorithm i in terms of Propositional Logic (PL) sentences α and β , and $\models$ and $\vdash_i$.

★ **SOLUTION:** An inference algorithm *i* is sound if whenever $KB \vdash_i \alpha$, it is also true that $KB \models \alpha$.

5. [20] Deductive Inference in Propositional Logic

Consider the following logic puzzle, one of many created by Lewis Carroll, the author of Alice in Wonderland:

```
No birds, except ostriches, are 9 feet high.
There are no birds in this aviary that belong to anyone but me.
No ostrich lives on mince pies.
I have no birds less than 9 feet high.
```

In this problem, you will use Resolution to prove that these premises imply the following conclusion:

```
Any bird in this aviary does not live on mince pies.
```

Consider the following propositional logic knowledge base (KB) that encodes these premises:

$$Bird \wedge Tall \Rightarrow Ostrich$$
$$Ostrich \Rightarrow Bird$$
$$Bird \Rightarrow IsMine$$
$$Ostrich \Rightarrow \neg LivesOnPies$$
$$IsMine \wedge Bird \Rightarrow Tall$$

To prove: $Bird \Rightarrow \neg LivesOnPies$

(PROBLEM ITSELF ON NEXT PAGE)

(Abbreviating in solution:

B - Bird
T – Tall
O – Ostrich
M – IsMine
L – LivesOnPies)

A. [10 points] Convert the premise sentences into conjunctive normal form (CNF) and show the result as a set of **clauses**.

★ **SOLUTION**: Each propositional logic sentence turns into one CNF clause as follows:

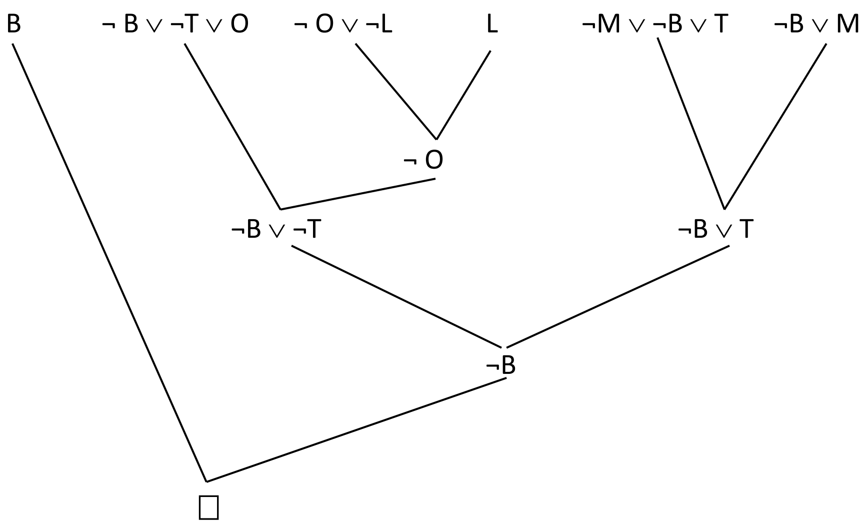
$(B \wedge T) \Rightarrow O$	$\neg B \vee \neg T \vee O$
$O \Rightarrow B$	$\neg O \vee B$
$B \Rightarrow M$	$\neg B \vee M$
$O \Rightarrow \neg L$	$\neg O \vee \neg L$
$(M \wedge B) \Rightarrow T$	$\neg M \vee \neg B \vee T$

To prove: $B \Rightarrow \neg L$ $\quad \neg B \vee L$ Negated, this turns into two clauses: $B \quad L$

B. [10 points] Now use resolution to show that KB $\models$ *Bird* $\Rightarrow$ $\neg$ *LivesOnPies* through proof by contradiction.

★ **SOLUTION**: There are many resolution proofs that are equivalent. Here's one:

In graphical form



One way to write this out:

$\neg O \vee \neg L$	and	L	resolve, yielding	$\neg O$
$\neg M \vee \neg B \vee T$	and	$\neg B \vee M$	resolve, yielding	$\neg B \vee T$
$\neg B \vee \neg T \vee O$	and	$\neg O$	resolve, yielding	$\neg B \vee \neg T$
$\neg B \vee \neg T$	and	$\neg B \vee T$	resolve, yielding	$\neg B$
B	and	$\neg B$	resolve, yielding	$\square$, showing the original is unsatisfiable