UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
CS440/ECE448 Artificial Intelligence
Conflict Exam 1
Spring 2026

Main Exam is February 16, 2026

Your Name: ______________________________

Your NetID: ______________________________

Instructions

- Please write your name on the top of every page.
- Have your ID ready; you will need to show it when you turn in your exam.
- This will be a CLOSED BOOK, CLOSED NOTES exam. You are permitted to bring and use only one 8.5x11 page of notes, front and back, handwritten or typed in a font size comparable to handwriting.
- No electronic devices (phones, tablets, calculators, computers etc.) are allowed.
- Make sure that your answer includes only the variables that it should include, but DO NOT simplify explicit numerical expressions. For example, the answer $x = \frac{1}{1+\exp(-0.1)}$ is MUCH preferred (much easier for us to grade) than the answer $x = 0.524979$.

Possibly Useful Formulas

$$P(X=x|Y=y)P(Y=y) = P(Y=y|X=x)P(X=x)$$
$$P(X=x) = \sum_y P(X=x,Y=y)$$
$$E\left[f(X,Y)\right] = \sum_{x,y} f(x,y)P(X=x,Y=y)$$

Precision,Recall $= \frac{TP}{TP+FP}, \frac{TP}{TP+FN}$

MPE=MAP: $f(x) = \arg\max\left(\log P(Y=y) + \log P(X=x|Y=y)\right)$

Naive Bayes: $P(X=x|Y=y) \approx \prod_{i=1}^{n} P(W=w_i|Y=y)$

Laplace Smoothing w/OOV: $P(W=w_i) = \frac{k+\text{Count}(W=w_i)}{k+\sum_v (k+\text{Count}(W=v))}$

Fairness: $P(Y|A) = \frac{P(Y|f(X),A)P(f(X)|A)}{P(f(X)|Y,A)}$

Score: $v_{t+1}(j) = \max_i v_t(i) + \log a_{ij} + \log b_j(x_{t+1})$

Backpointer: $\psi_{t+1}(j) = \arg\max_i v_t(i) + \log a_{ij} + \log b_j(x_{t+1})$

Linear Regression: $\mathscr{L} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{w}^T\mathbf{x}_i - y_i\right)^2$

Question 1 ***(26 points)***

Lochs, lagoons, and lakes are all bodies of water, but they tend to be inhabited by different types of supernatural creatures. Let S be the type of supernatural creature in a body of water (1 =dinosaur, 2 =merman, 3 =fairy), and let L be the class label (1 =loch, 2 =lagoon, 3 =lake). Define $P(L=\ell)$ to be the prior probability of class ℓ, and $P(S=s|L=\ell)$ to be the likelihood of creature s in setting ℓ.

(a) (13 points) A Bayes classifier observes m, then chooses the value of ℓ with the highest *a posteriori* probability. In terms of the distributions $P(L)$ and $P(S|L)$, what is the error rate of the Bayes classifier? Your answer may contain explicit summations, minimizations, and/or maximizations; please specify the variable(s) over which you are summing, maximizing, and/or minimizing.

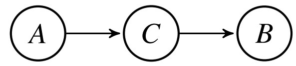
> **Solution:**
>
> $$P(\text{Error}) = 1 - \sum_s \max_\ell P(L=\ell)P(S=s|L=\ell)$$

(b) (13 points) The famous explorer Ibn Battuta visited 5 lochs. He found that 3 lochs were inhabited by dinosaurs, while 2 were inhabited by mermen. Assuming that every body of water is inhabited by exactly one supernatural creature of class $s \in \{1,2,3\}$ (assume that exceptions to this rule are impossible), in terms of a Laplacian smoothing parameter, estimate $P(S=s|L=1)$ for all $s \in \{1,2,3\}$.

> **Solution:**
>
> $$P(S=s|L=1) = \begin{cases} \frac{3+k}{5+3k} & s=1 \\ \frac{2+k}{5+3k} & s=2 \\ \frac{k}{5+3k} & s=3 \end{cases}$$

Question 2 ***(30 points)***

Consider three binary events A, B, and C, related by the Bayes network shown here:



The parameters of this Bayes network are given by the unknown constants v through z, as follows:

$$P(A) = v$$
$$P(C|A) = w, \quad P(C|\neg A) = x$$
$$P(B|C) = y, \quad P(B|\neg C) = z$$

(a) (10 points) Are events A and B independent, conditionally independent given knowledge of C, both, or neither? Explain your answer.

Solution: They are not independent, because they have an unknown common ancestor (A is an ancestor of both). They are conditionally independent given knowledge of C, because knowledge of C cuts the ancestor connection between A and B, so they no longer have any unknown common ancestor or known common descendent.

(b) (10 points) Suppose C is unknown. Find the four probabilities $P(A,B)$, $P(A,\neg B)$, $P(\neg A,B)$, and $P(\neg A,\neg B)$.

Solution:

$$P(A,B) = vwy + v(1-w)z$$
$$P(A,\neg B) = vw(1-y) + v(1-w)(1-z)$$
$$P(\neg A,B) = (1-v)xy + (1-v)(1-x)z$$
$$P(\neg A,\neg B) = (1-v)x(1-y) + (1-v)(1-x)(1-z)$$

(c) (10 points) Suppose C is known to be true. Find $P(C)$, and in terms of $P(C)$ find $P(A,B|C)$, $P(A,\neg B|C)$, $P(\neg A,B|C)$, and $P(\neg A,\neg B|C)$.

Solution:

$$P(C) = vwy + vw(1-y) + (1-v)xy + (1-v)x(1-y) = vw + (1-v)x$$

$$P(A,B|C) = \frac{vwy}{P(C)}$$

$$P(A,\neg B|C) = \frac{vw(1-y)}{P(C)}$$

$$P(\neg A,B|C) = \frac{(1-v)xy}{P(C)}$$

$$P(\neg A,\neg B|C) = \frac{(1-v)x(1-y)}{P(C)}$$

Question 3 ***(20 points)***

Consider an HMM with binary state variable Y_t, binary observation variable X_t, transition probabilities $a_{ij} = P(Y_{t+1} = j|Y_t = i)$, and observation probabilities $b_{ij} = P(X_t = j|Y_y = i)$, and suppose you know that $Y_0 = 0$. Suppose there is some normalizer, z, and base, c, such that $\log_c(a_{ij}/z) = \log_c(b_{ij}/z) = 0$ for $i = j$, but for $i \neq j$, $\log_c(b_{ij}/z) = -2$ and $\log_c(a_{ij}/z) = -3$. Suppose that $\{X_1, \ldots, X_9\} = \{1, 0, 0, 1, 1, 1, 0, 0, 1\}$. Draw a trellis specifying the numerical values of $v_t(i) = \max_{Y_\tau, \tau<t} \log_c P(Y_1, \ldots, Y_t = i|X_1, \ldots, X_t)/z$ for $i \in \{0, 1\}$ and $1 \leq t \leq 9$, and specify a sequence of state variables $\{Y_1, \ldots, Y_9\}$ that maximizes the score. If there are more than one state sequences with the same maximum score, you only need to provide one of them.

Solution: A correct answer needs to show the following numerical values. The arrows need not be shown, but may be helpful to you in solving the second part of the question:

t	1	2	3	4	5	6	7	8	9
X_t	1	0	0	1	1	1	0	0	1
$v_t(0)$	$\leftarrow -2$	$\leftarrow -2$	$\leftarrow -2$	$\leftarrow -4$	$\leftarrow -6$	$\leftarrow -8$	-8	$\leftarrow -8$	$\leftarrow -10$
$v_t(1)$	$\nwarrow -3$	-5	-7	$\nwarrow -5$	$\leftarrow -5$	$\leftarrow -5$	$\leftarrow -7$	$\leftarrow -9$	$\leftarrow -9$

The maximum-likelihood state sequence is $\{0, 0, 0, 1, 1, 1, 1, 1, 1\}$.

Question 4 ***(24 points)***

Consider a linear regression model $f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i$ where $\mathbf{x}_i$ is an m-dimensional vector drawn from a size-n training corpus, $1 \le i \le n$. Suppose you want to choose the vector $\mathbf{w}$ to minimize $\mathscr{L}_{\text{train}}$, defined as

$$\mathscr{L}_{\text{train}} = \frac{1}{n}\sum_{i=1}^{n} |f(\mathrm{x}_i) - y_i|^5,$$

where y_i is a real-valued scalar for $1 \le i \le n$.

(a) (12 points) Find the gradient $\frac{\partial \mathscr{L}_{\text{train}}}{\partial \mathbf{w}}$.

Solution:

$$\frac{\partial \mathscr{L}_{\text{train}}}{\partial \mathbf{w}} = \frac{1}{n}\sum_{i=1}^{n} 5\mathbf{x}_i\,|f(\mathbf{x}_i) - y_i|^4 \operatorname{sign}\left(f(\mathbf{x}_i) - y_i\right)$$

(b) (12 points) Suppose you also have a development corpus, with tokens numbered $n+1 \le i \le n+j$, and a test corpus, with tokens numbered $n+j+1 \le i \le n+j+k$. The losses in these corpora are

$$\mathscr{L}_{\text{dev}} = \frac{1}{j}\sum_{i=n+1}^{n+j} |f(\mathrm{x}_i) - y_i|^5,$$

$$\mathscr{L}_{\text{test}} = \frac{1}{k}\sum_{i=n+j+1}^{n+j+k} |f(\mathrm{x}_i) - y_i|^5.$$

Your client, a hardware manufacturer, is very clever about finding things to measure. They propose the following experiment: (1) increase m by finding something else to measure about each product, (2) re-train the new $\mathbf{w}$, finding the value that minimizes $\mathscr{L}_{\text{train}}$, (3) measure $\mathscr{L}_{\text{dev}}$ and $\mathscr{L}_{\text{test}}$, (4) repeat.

Do you expect $\mathscr{L}_{\text{test}}$ to be smallest when $m = 1$, when $m = \frac{n}{2}$, or when $m = \frac{3n}{2}$? Why?

Solution: $\mathscr{L}_{\text{test}}$ is the sum of optimization error plus generalization error. Optimization error gets smaller as m increases, but generalization error gets larger as m increases, therefore I would expect $m = \frac{n}{2}$ to have the smallest $\mathscr{L}_{\text{test}}$.

This page is scratch paper