UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CS440/ECE448 Artificial Intelligence

Practice Exam 1

Spring 2026

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

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

$$\textbf{Precision,Recall} \quad = \frac{TP}{TP+FP}, \frac{TP}{TP+FN}$$
$$\textbf{MPE=MAP:} \quad f(x) = \arg\max\left(\log P(Y=y) + \log P(X=x|Y=y)\right)$$

$$\textbf{Naive Bayes:} \quad P(X=x|Y=y) \approx \prod_{i=1}^{n} P(W=w_i|Y=y)$$
$$\textbf{Laplace Smoothing w/OOV:} \quad P(W=w_i) = \frac{k+\text{Count}(W=w_i)}{k+\sum_v (k+\text{Count}(W=v))}$$
$$\textbf{Fairness:} \quad P(Y|A) = \frac{P(Y|f(X),A)P(f(X)|A)}{P(f(X)|Y,A)}$$

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

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

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

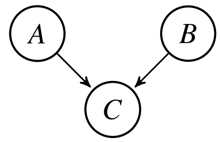
You studied well past midnight, so when you wake up, you're not sure what time it is. You estimate that it's morning with probability P_{AM}, and afternoon with probability $P_{PM} = 1 - P_{AM}$. Your cafeteria serves only one type of pastry at a time; pastries in the morning have fruit with probability $P(F|AM)$ and meat with probability $P(M|AM) = 1 - P(F|AM)$, while the corresponding likelihoods for pastries in the afternoon are $P(M|PM) = 1 - P(F|PM)$.

(a) (13 points) A Bayes classifier orders a pastry, then computes the time of day that has the highest *a posteriori* probability given the type of pastry. In terms of the parameters P_{AM}, P_{PM}, $P(F|AM)$, $P(F|PM)$, $P(M|AM)$, and/or $P(M|PM)$, what is the error rate of such a 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.

(b) (13 points) In the past 30 days you've awakened before noon 25 times, after noon 5 times. You received a fruit pastry before noon 16 times & after noon once; you received a meat pastry before noon 9 times and after noon 4 times. Assuming that all pastries have either meat or fruit but not both (any other option is impossible), and in terms of a Laplace smoothing parameter k, estimate P_{AM}, P_{PM}, $P(F|AM)$, $P(M|AM)$, $P(F|AM)$, and $P(F|PM)$.

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 u through z, as follows:

$$\begin{aligned} P(A) = u, &\quad P(B) = v \\ P(C|A,B) = w, &\quad P(C|A,\neg B) = x \\ P(C|\neg A,B) = y, &\quad P(C|\neg A,\neg B) = z \end{aligned}$$

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

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

(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 $P(\neg A,\neg B|C)$.

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

You've been hired as a night watchman at a nuclear power plant. Let $Y_t = 1$ if the power plant is overheating at time t, otherwise $Y_t = 0$, and suppose that $Y_0 = 0$. Let $X_t = 1$ if the reactor warning light is blinking at time t, otherwise $X_t = 0$. Define $a_{ij} = P(Y_{t+1} = j|Y_t = i)$, and $b_{ij} = P(X_t = j|Y_y = i)$. 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 $i \neq j$, $\log_c(a_{ij}/z) = -2$ and $\log_c(b_{ij}/z) = -3$. Suppose that $\{X_1,\ldots,X_9\} = \{1,0,0,1,1,1,0,0,1\}$. Draw a trellis specifying the numerical values of $v_t(i) = \max \log_c P(Y_1,\ldots,Y_t = i|X_1,\ldots,X_t)/z$ 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.

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

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

where ln is the natural logarithm, and y_i is a real-valued scalar for $1 \le i \le n$.

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

(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

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

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

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

This page is scratch paper