

Lecture 17: Exam 1 Review

Mark Hasegawa-Johnson, 2/2024

These slides are public domain: Copy at will

Outline

- How to take the exam
- What's on the exam
- Sample problems

How to take the exam

- Go to <https://cbtf.illinois.edu>, choose CS440/ECE448 exam 1, and register the time and location where you will take the exam
- If you can't register at cbtf.illinois.edu, contact us on campuswire BEFORE 1pm Monday

How to take the exam

- Show up on time at the appointed location! It's possible to reschedule if you miss an exam, but it will require you to go to a CBTF location to apply in person for permission to do so.
- Bring: Pencils and erasers
- CBTF will provide: Scratch paper
- The exam will have attached: a PDF formula sheet, which is also available now on the course web page, so you can see what will be on it

Exam format

- The exam will have 8 questions, each worth 13 points --- 104 points total. If you get 104 out of 104, that is worth 100%.
 - $100/104 = 96\%$.
- Each question will be multiple choice, very similar to the daily quiz questions in class
- You will get full credit (13 points) if you get it right on the first try
- If not, try again (and again and again and again, if necessary), for reduced credit
 - E.g., if a question has 6 options, then you get 13,9,7,5,3,1 points for getting the correct answer on the 1st/2nd/3rd/4th/5th/6th try.

Exam format

- Formula sheet
- Questions to answer

The screenshot shows a web browser window with the URL https://us.prairielearn.com/pl/course_instance/147925/assessment_insta.... The page title is "E1: Exam 1". The user is logged in as "Mark Hasegawa-Johnson" with a student ID of 17. The assessment is for "ECE 448, sp2024".

Key information on the page:

- Total points: 104/104 (100%)
- Assessment is **open** and you can answer questions.
- Available credit: 100% (Staff override)
- A link for the [Formula sheet](#) is provided.

Question	Submission status	Best submission ?	Available points ?	Awarded points ?
Decision Theory, Naive Bayes, Bayesian Networks and HMMs				
Question 1	complete	100%	—	13 /13
Question 2	complete	100%	—	13 /13
Question 3	complete	100%	—	13 /13
Question 4	complete	100%	—	13 /13
Question 5	complete	100%	—	13 /13
Learning and Neural Networks				
Question 6	complete	100%	—	13 /13
Question 7	complete	100%	—	13 /13
Question 8	complete	100%	—	13 /13

At the bottom, there is a message: "No saved answers to grade". Below this, there are instructions:

- Submit your answer to each question with the **Save & Grade** or **Save only** buttons on the question page.
- Look at **Best submission** to confirm that each question has been graded. Questions with **Available points** can be attempted again for more points. Attempting questions again will never reduce the points you already have.

Outline

- How to take the exam
- What's on the exam
- Sample problems

Topics

- Bayesian networks – 5 questions
 - probability, decision theory, naïve Bayes, Bayes nets, HMMs, fairness
- Neural networks – 3 questions
 - learning, regression, perceptron, softmax, multilayer, vision, CNN

Probability

- Probability Mass and Probability Density

$$P(X = x) = \Pr(X = x) \quad \dots \quad \text{or} \quad \dots \quad P(X = x) = \frac{d}{dx} \Pr(X \leq x)$$

- Jointly random variables

$$P(X = \mathbf{x}) = P(X_1 = x_1, \dots, X_n = x_n)$$

- Conditional Probability and Independence

$$P(X, Y) = P(X|Y)P(Y)$$
$$P(X|Y) = P(X) \Leftrightarrow P(X, Y) = P(X)P(Y)$$

- Expectation

$$E[f(X, Y)] = \sum_{x, y} f(x, y)P(X = x, Y = y) \quad \dots \quad \text{or} \quad \dots$$
$$E[f(X, Y)] = \iint_{-\infty}^{\infty} f(x, y)P(X = x, Y = y)dxdy$$

- Mean, Variance and Covariance

$$\Sigma = E[(X - E[X])(X - E[X])^T]$$

Decision theory

- Minimum Probability of Error = Maximum *A Posteriori*:

$$f(x) = \operatorname{argmax}_y P(Y = y|X = x)$$

- Bayes Error Rate:

$$\text{Bayes Error Rate} = \sum_x P(X = x) \min_y P(Y \neq y|X = x)$$

- Confusion Matrix, Precision & Recall, Sensitivity & Selectivity

$$\text{Precision} = P(Y = 1|f(X) = 1) = \frac{TP}{TP + FP}$$

$$\text{Sensitivity} = \text{Recall} = P(f(X) = 1|Y = 1) = \frac{TP}{TP + FN}$$

$$\text{Selectivity} = P(f(X) = 0|Y = 0) = \frac{TN}{TN + FP}$$

- Train, Dev, and Test Corpora

Naïve Bayes

- MPE = MAP with Bayes' rule:

$$f(x) = \operatorname{argmax}_y (\log P(Y = y) + \log P(X = x|Y = y))$$

- naïve Bayes:

$$\log P(X = x|Y = y) \approx \sum_{i=1}^n \log P(W = w_i|Y = y)$$

- maximum likelihood parameter estimation:

$$P(W = w_i|Y = y) = \frac{\operatorname{Count}(w_i, y)}{\sum_{v \in V} \operatorname{Count}(v, y)}$$

- Laplace Smoothing:

$$P(W = w_i|Y = y) = \begin{cases} \frac{k + \operatorname{Count}(w_i, y)}{k + \sum_{v \in V} (k + \operatorname{Count}(v, y))} & W = \text{OOV is possible} \\ \frac{k + \operatorname{Count}(w_i, y)}{\sum_{v \in V} (k + \operatorname{Count}(v, y))} & \text{otherwise} \end{cases}$$

Bayesian Networks

- Bayesian classifier: $f(x) = \underset{y}{\operatorname{argmax}} P(Y = y|X = x)$
- Bayesian network: A better way to represent knowledge
 - Each variable is a node.
 - An arrow between two nodes means that the child depends on the parent.
- Inference using a Bayesian network

$$P(B = \top, J = \top) = \sum_{e=\top}^{\perp} \sum_{a=\top}^{\perp} P(B = \top)P(E = e)P(A = a|B = \top, E = e)P(J = \top|A = a)$$

- Key ideas: Independence and Conditional independence
 - Independent = no common ancestors
 - Conditionally independent = (1) no common descendants, and (2) none of the descendants of one are ancestors of the other

HMM: Viterbi Algorithm

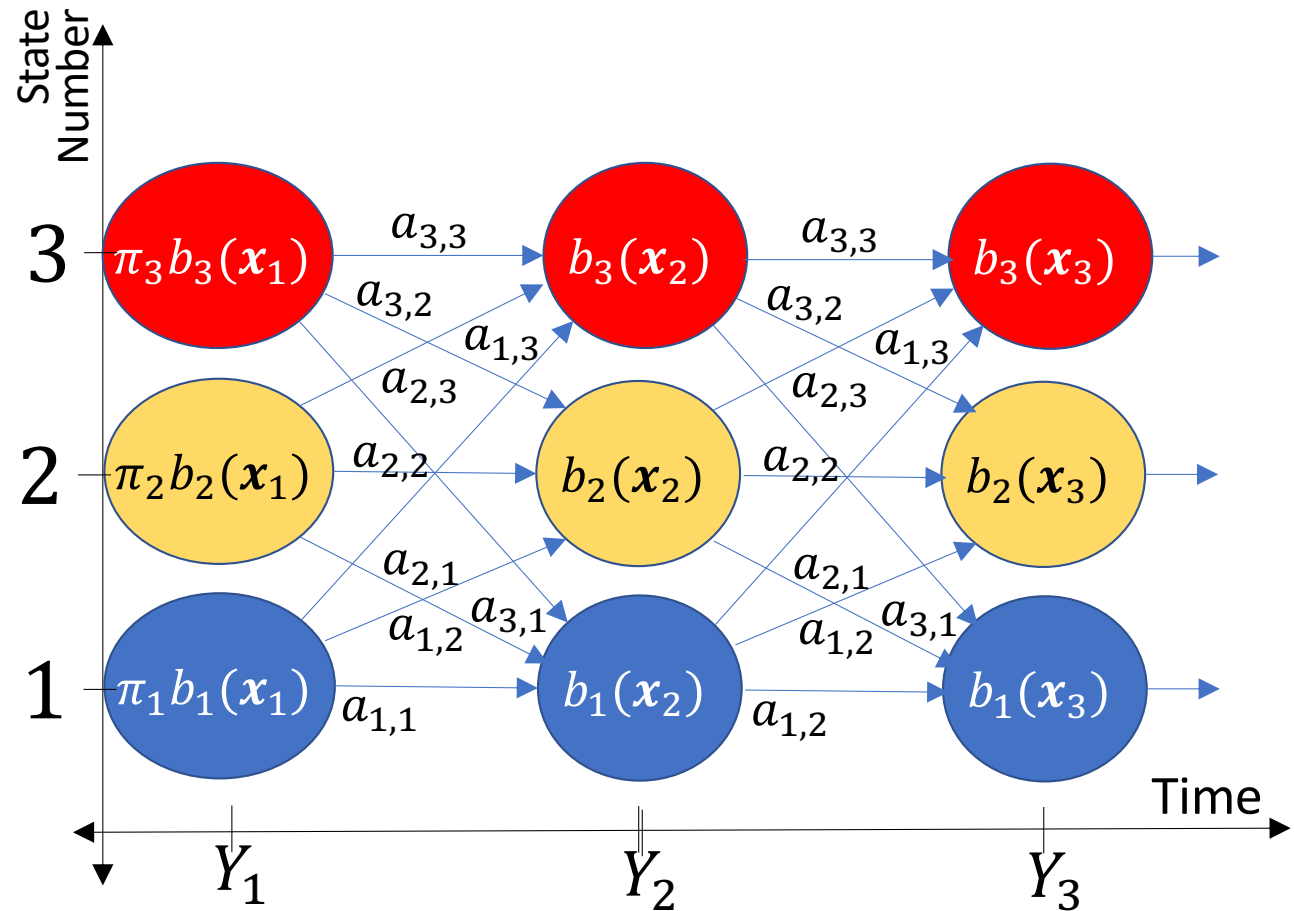
Trellis is used to find the most likely path.

For example:

$$v_t(j) = \max_i v_{t-1}(i) a_{i,j} b_j(\mathbf{x}_t)$$

$$\psi_t(j)$$

$$= \operatorname{argmax}_i v_{t-1}(i) a_{i,j} b_j(\mathbf{x}_t)$$



Three Definitions of Fairness

- Demographic Parity: $P(f(X)|A = 1) = P(f(X)|A = 0)$
- Equal Odds: $P(f(X)|Y, A = 1) = P(f(X)|Y, A = 0)$
- Predictive Parity: $P(Y|f(X), A = 1) = P(Y|f(X), A = 0)$

Those three things can only all be true, all at the same time, if:

- $P(Y|A = 1) = P(Y|A = 0)$

Learning

- **Biological inspiration:** Neurons that fire together wire together. Given enough training examples (x_i, y_i) , can we learn a desired function so that $f(x) \approx y$?
- **Classification tree:** Learn a sequence of if-then statements that computes $f(x) \approx y$
- **Mathematical definition of supervised learning:** Given a training dataset, $\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\}$, find a function f that minimizes the risk, $\mathcal{R} = \mathbb{E}[\ell(Y, f(X))]$.
- **Overtraining:** $\mathcal{R}_{\text{emp}} = \frac{1}{n} \sum_{i=1}^n \ell(y_i, f(x_i))$ reaches zero if you train long enough.
- **Early Stopping:** Stop when error rate on the dev set reaches a minimum

Linear Regression

- Definition of linear regression

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

- Mean-squared error

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n \mathcal{L}_i, \quad \mathcal{L}_i = \frac{1}{2} \epsilon_i^2, \quad \epsilon_i = f(\mathbf{x}_i) - y_i$$

- Gradient descent

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}}, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{1}{n} \sum_{i=1}^n \epsilon_i \mathbf{x}_i$$

- Stochastic gradient descent

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \frac{\partial \mathcal{L}_i}{\partial \mathbf{w}}, \quad \frac{\partial \mathcal{L}_i}{\partial \mathbf{w}} = \epsilon_i \mathbf{x}_i$$

Perceptron

- Linear Classifiers: $f(\mathbf{x}) = \operatorname{argmax} \mathbf{W}\mathbf{x} + \mathbf{b}$

- One-hot vectors: $\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_v(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbb{1}_{\operatorname{argmax} \mathbf{W}\mathbf{x}=1} \\ \vdots \\ \mathbb{1}_{\operatorname{argmax} \mathbf{W}\mathbf{x}=v} \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbb{1}_{y=1} \\ \mathbb{1}_{y=2} \\ \vdots \end{bmatrix}$

- Perceptron learning algorithm:

$$\mathbf{w}_c \leftarrow \begin{cases} \mathbf{w}_c - \eta \mathbf{x} & c = \operatorname{argmax} \mathbf{W}\mathbf{x} + \mathbf{b} \\ \mathbf{w}_c + \eta \mathbf{x} & c = y \\ \mathbf{w}_c & \text{otherwise} \end{cases}$$

- Softmax: $f_c(\mathbf{x}) = \frac{\exp(\mathbf{w}_c^T \mathbf{x} + b_c)}{\sum_{k=1}^v \exp(\mathbf{w}_k^T \mathbf{x} + b_k)} \approx \Pr(Y = c | \mathbf{x})$

- Sigmoid: $\sigma(\mathbf{w}^T \mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x} + b)}} \approx \Pr(Y = 1 | \mathbf{x})$

- Cross-entropy: $\mathcal{L} = -\ln f_y(\mathbf{x})$, $\frac{\partial \mathcal{L}}{\partial f_c(\mathbf{x})} = \begin{cases} -\frac{1}{f_c(\mathbf{x})} & c = y \\ 0 & \text{otherwise} \end{cases}$

- Gradient descent: $\mathbf{w}_c \leftarrow \mathbf{w}_c - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}_c}$

- Derivative of the cross-entropy of a softmax:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_c} = \epsilon_c \mathbf{x}, \quad \epsilon_c = \begin{cases} f_c(\mathbf{x}_i) - 1 & c = y \text{ (output should be 1)} \\ f_c(\mathbf{x}_i) - 0 & \text{otherwise (output should be 0)} \end{cases}$$

Multi-Layer

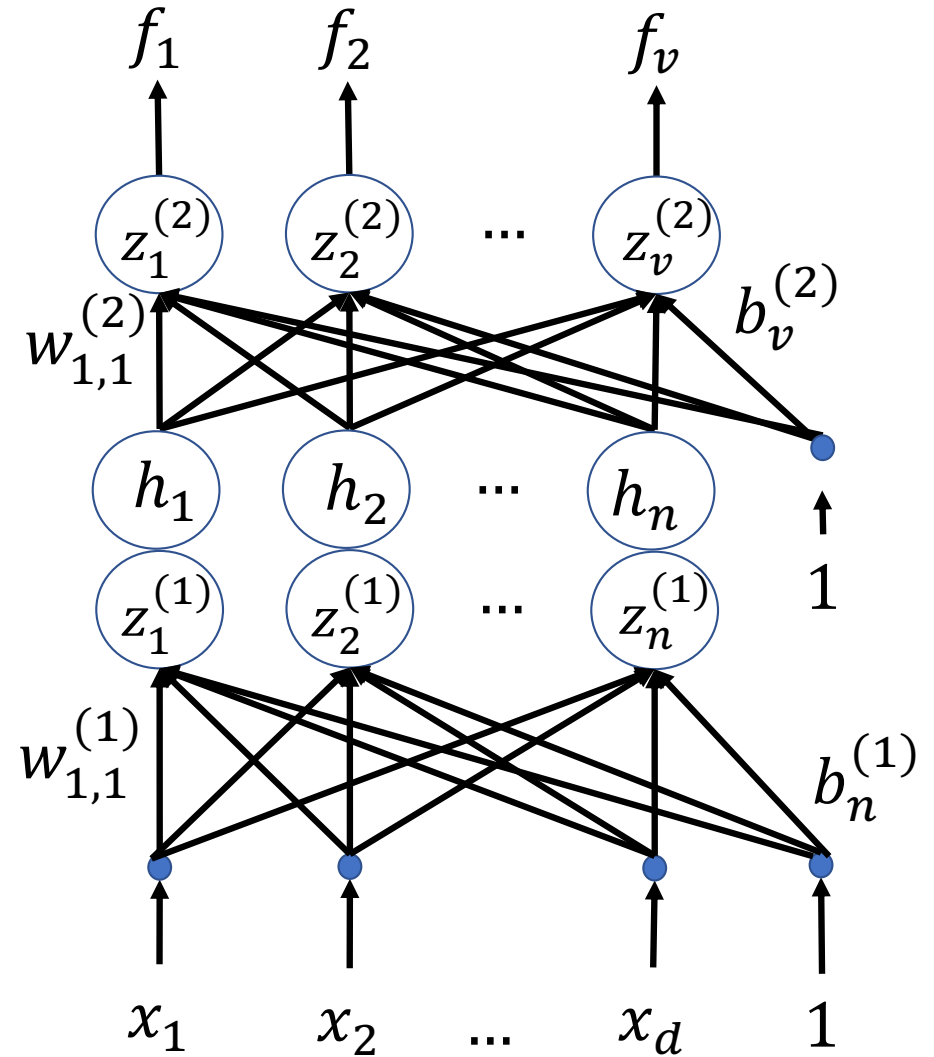
For example:

$$f_k = \underset{k}{\text{softmax}} \mathbf{z}^{(2)}$$

$$z_k^{(2)} = b_k^{(2)} + \sum_{j=1}^n w_{k,j}^{(2)} h_j$$

$$h_j = \text{ReLU}(z_j^{(1)})$$

$$z_j^{(1)} = b_j^{(1)} + \sum_{i=1}^d w_{j,i}^{(1)} x_i$$



Back-propagation

For example, if the loss is cross-entropy, then

$$\frac{\partial \mathcal{L}}{\partial z_k^{(2)}} = f_k - \mathbb{1}_{y=k}$$

So the weight gradient is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_{j,i}^{(1)}} &= \sum_{k=1}^v \left(\frac{\partial \mathcal{L}}{\partial z_k^{(2)}} \right) \left(\frac{\partial z_k^{(2)}}{\partial h_j} \right) \left(\frac{\partial h_j}{\partial w_{j,i}^{(1)}} \right) \\ &= \sum_{k=1}^v (f_k - \mathbb{1}_{y=k}) w_{k,j}^{(2)} \mathbb{1}_{h_j > 0} x_i \end{aligned}$$

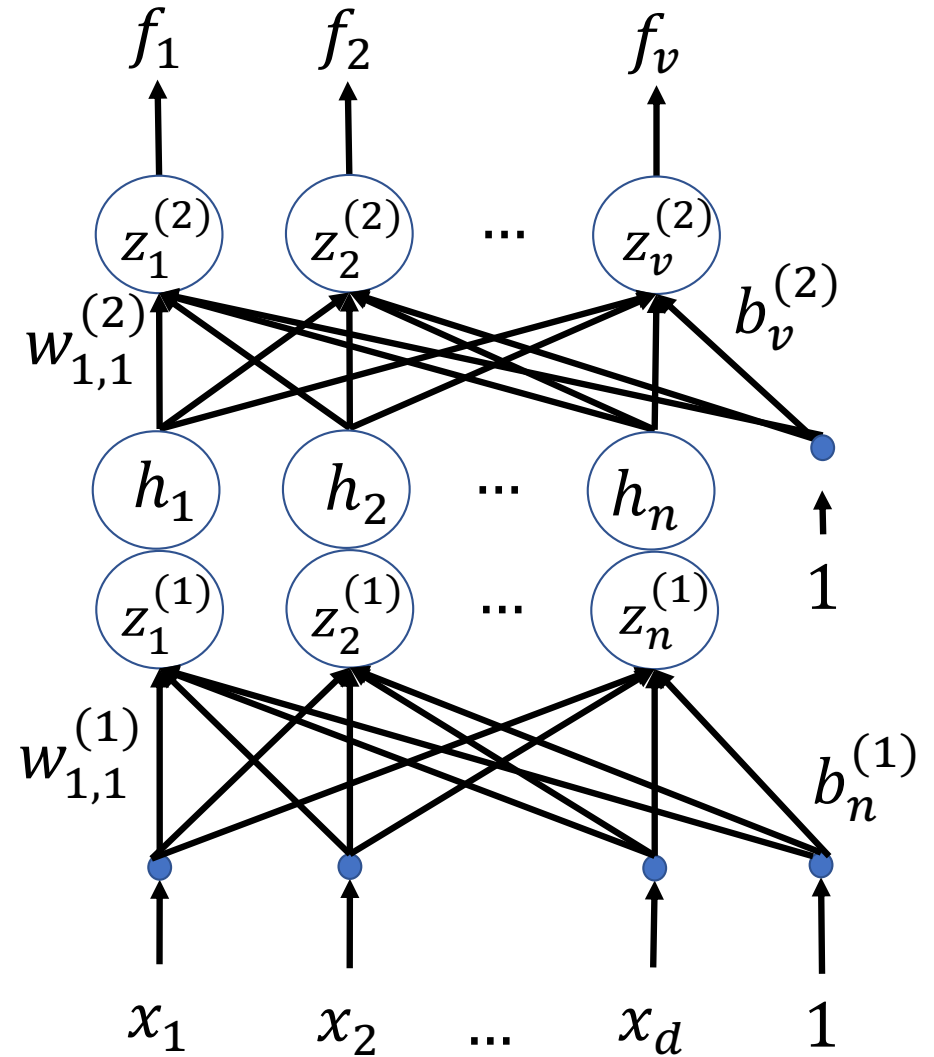


Image formation & processing

- Pinhole camera equations:

$$\frac{x'}{f} = -\frac{x}{z}, \quad \frac{y'}{f} = -\frac{y}{z}$$

- Vanishing point = parallel lines:

$$\begin{aligned} x_1 &= az + c_1, & y_1 &= bz + d_1 \\ x_2 &= az + c_2, & y_2 &= bz + d_2 \end{aligned}$$

- Edge detection using difference-of-Gaussians:

$$h(m, n) = \frac{1}{2\pi\sigma^2} e^{-\left(\left(\frac{m}{\sigma}\right)^2 + \left(\frac{n}{\sigma}\right)^2\right)}$$

$$h_x'(x', y') = \frac{(h(x' + 1, y') - h(x' - 1, y'))}{2}$$

$$h_y'(x', y') = \frac{(h(x', y' + 1) - h(x', y' - 1))}{2}$$

Convolution and Max Pooling

$$y[k, l] = h[k, l] * x[k, l] = \sum_i \sum_j x[k - i, l - j] h[i, j]$$

$$\frac{d\mathcal{L}}{dh[i, j]} = \sum_k \sum_l \frac{d\mathcal{L}}{dy[k, l]} \frac{dy[k, l]}{dh[i, j]}$$

$$z[m, n] = \max_{\substack{(m-1)p+1 \leq k \leq mp, \\ (n-1)p+1 \leq l \leq np}} y[k, l]$$

$$\frac{d\mathcal{L}}{dy[k, l]} = \begin{cases} \frac{d\mathcal{L}}{dz[m, n]} & \text{if } y[k, l] = \max_{\substack{(m-1)p+1 \leq i \leq mp, \\ (n-1)p+1 \leq j \leq np}} y[i, j] \\ 0 & \text{otherwise} \end{cases}$$

Outline

- How to take the exam
- What's on the exam
- **Sample problems**

Sample problems

- You can do the daily quizzes again, as often as you like
 - Only your highest score counts toward your grade (presumably this is the one you received the first time you did the quiz, when 100% was possible)
 - After you get a problem correct, you can click “Try a new variant” to try a new variant
- Sample exam is available
 - Coverage is similar to the exam next week
 - Style of questions is quite different

Sample problems

- Sample problems
 - NOT the same format as the real exam
- Formula sheet
 - SAME as the real exam

speech - Search Results - Publi X clientFiles and serverFiles - Pr X Exams - CS440/ECE448 Artificial I X

https://courses.grainger.illinois.edu/ece448/sp2024/exam 110%

CS440/ECE448 Artificial Intelligence » Grades » Exams previous | next | index

Exams

Exams this semester will be done at CBTF.

Exams will be closed-book, closed-notes, no calculator. Each exam will have a title page with formulas on it, but you will not be able to bring your own formula sheet to the exam.

Exam questions will be multiple-choice, like the in-class quizzes. If you miss a question the first time, you will be permitted to try another choice, for a reduced point total.

Conflict exams are available if you are traveling or otherwise unavailable during the three days that we offer an exam. If you know that you have such a conflict, please e-mail the instructor no later than 7 days prior to the exam date. Conflict exams will take place at CTBF.

Exam 1

Exam 1 will be held at CBTF during the window M-W, February 26-28. It will be composed of eight multiple-choice questions. Each question will be similar to one of the daily quizzes. You will be able to try all of the available answers if you wish, but you will receive fewer points for a correct answer on the N'th attempt than for a correct answer on the (N-1)'st attempt.

- This [sample exam](#) covers the material that will be on the exam, and here are its [solutions](#).
- This [formula sheet](#) will be provided on the PrairieTest exam.

Exam 2

Exam 2 will be held at CBTF during the window M-W, April 15-17.

Final Exam

Exam 3 will be held in three sessions. You will need to bring your own device on which to do the exam; if you need one, loaner laptops are available from Engineering IT. The three available sessions will be:

Table of Contents

- Exams
 - Exam 1
 - Exam 2
 - Final Exam

Previous topic
Quizzes, Machine Problems, and Lectures

Next topic
4-Credit Section

This Page
Show Source

Quick search
Go