

# Lecture 16: Exam 1 Review

Mark Hasegawa-Johnson

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

# 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.
- Most or all will be questions that you have seen in daily quizzes. Most are multiple choice, though some are short-answer.
- Points per question: 10,8,6,4,2 if you get it right on the first, second, third, fourth, fifth try.

# Exam format

- Formula sheet
- Questions to answer

Course staff: [Regenerate your assessment instance](#) to pick up changes to the assessment or to get a fresh set of questions.

E1: Exam 1

Total points: 0/80 0% Assessment is **open** and you can answer questions. Available credit: 100% (Staff override)

[Formula sheet](#)

Question	Status	Available points	Awarded points
<b>Exam 1</b>			
<a href="#">Question 1</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 2</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 3</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 4</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 5</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 6</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 7</a>	unanswered	10, 8, 6, 4, 2	— /10
<a href="#">Question 8</a>	unanswered	10, 8, 6, 4, 2	— /10

No saved answers to grade

- Submit your answer to each question with the **Save & Grade** or **Save only** buttons on the question page.
- Look at **Status** 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

- Probability
- Decision theory and fairness
- Naïve bayes
- Learning and linear regression
- Logistic regression and word2vec
- Nonlinear regression and relevance backprop
- Softmax and transformer
- HMMs and Bayesian networks



# Probability

- Axioms of probability:

$$P(A) \geq 0, P(\Omega) = 1, P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$

- Jointly random variables

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

- Conditional Probability

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

- Independence

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

- Expectation

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

# Decision theory

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

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

- 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:

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

- Laplace Smoothing:

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

## 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

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

# Logistic Regression

- Logistic regression

$$f(\mathbf{x}_i) = \sigma(\mathbf{x}_i^T \mathbf{w}) = \frac{1}{1 + e^{-\mathbf{x}_i^T \mathbf{w}}}$$

- Derivative of the sigmoid

$$\frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1 - \sigma(z))$$

- Derivative of the negative log sigmoid

$$\mathcal{L} = - \sum_{y_i=1} \log \sigma(\mathbf{x}_i^T \mathbf{w}) - \sum_{y_i=0} \log(1 - \sigma(\mathbf{x}_i^T \mathbf{w}))$$

# word2vec

- Continuous bag of words:

$$\mathcal{L} = -\frac{1}{T} \sum_{t=0}^{T-1} \sum_{j=-c, j \neq 0}^c \ln P(w_t | w_{t+j})$$

- Skip-gram:

$$\mathcal{L} = -\frac{1}{T} \sum_{t=0}^{T-1} \sum_{j=-c, j \neq 0}^c \ln P(w_{t+j} | w_t)$$

- Dot-product similarity:

$$\mathcal{L} = -\frac{1}{T} \sum_{t=0}^{T-1} \sum_{j=-c, j \neq 0}^c \ln \sigma(\mathbf{v}_1^T \mathbf{v}_2)$$

- Skip-gram noise contrastive estimation:

$$\mathcal{L} = -\ln \sigma(\mathbf{v}_{t+j}^T \mathbf{v}_t) - \frac{1}{k} \sum_{i=1}^k \ln(1 - \sigma(\mathbf{v}_i^T \mathbf{v}_t))$$



# Nonlinear regression

- Piece-wise constant nonlinear regression:

$$f(\mathbf{x}) = \mathbf{w}^{(2),T} \sigma(\mathbf{W}^{(1)} \mathbf{x})$$

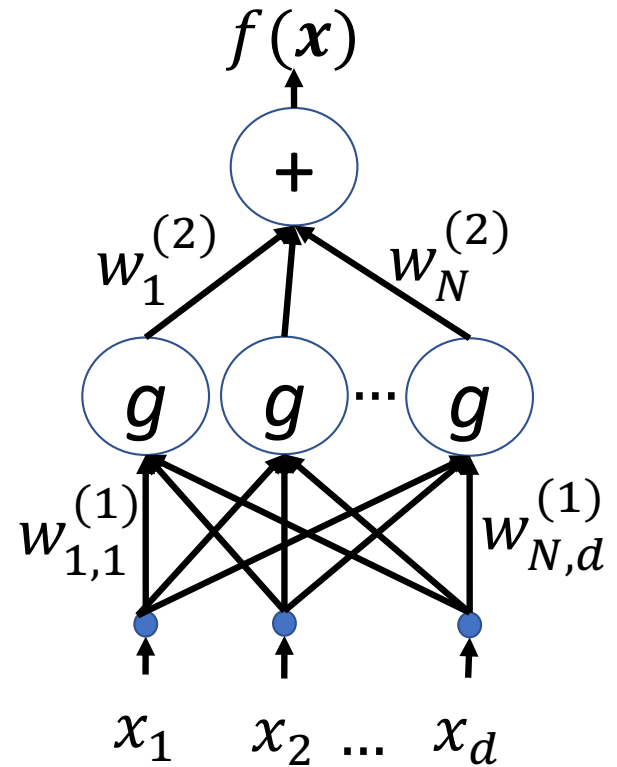
- Piece-wise linear regression:

$$f(\mathbf{x}) = \mathbf{w}^{(2),T} \text{ReLU}(\mathbf{W}^{(1)} \mathbf{x})$$

- Back-propagation:

$$\mathbf{W}^{(1)} \leftarrow \mathbf{W}^{(1)} - \eta \frac{\partial \mathcal{L}_i}{\partial \mathbf{W}^{(1)}}$$

$$\mathbf{w}^{(2)} \leftarrow \mathbf{w}^{(2)} - \eta \frac{\partial \mathcal{L}_i}{\partial \mathbf{w}^{(2)}}$$



# Softmax

- Multi-class linear classifiers

$$f(\mathbf{x}) = \operatorname{argmax} \mathbf{W}\mathbf{x}$$

- One-hot vectors

$$f_c(\mathbf{x}) = \mathbb{1}_{\operatorname{argmax} \mathbf{W}\mathbf{x}=c}$$

- Softmax nonlinearity

$$f_c(\mathbf{x}) = \frac{\exp(\mathbf{w}_c^T \mathbf{x})}{\sum_{k=1}^v \exp(\mathbf{w}_k^T \mathbf{x})}$$

- Derivative of the log softmax

$$\mathbf{w}_c \leftarrow \mathbf{w}_c + \eta \sum_{i=1}^n \left( \mathbb{1}_{y_i=c} - f_c(\mathbf{x}) \right) \mathbf{x}_i$$

# Transformer

- Query, Key, and Value matrices

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}, \mathbf{K} = \begin{bmatrix} \mathbf{k}_1^T \\ \vdots \\ \mathbf{k}_n^T \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_m^T \end{bmatrix}$$

- Attention

$$\mathbf{c}_i = \mathbf{V}^T \text{softmax}(\mathbf{K} \mathbf{q}_i) = \sum_t \frac{\exp(\mathbf{q}_i^T \mathbf{k}_t)}{\sum_\tau \exp(\mathbf{q}_i^T \mathbf{k}_\tau)} \mathbf{v}_t$$

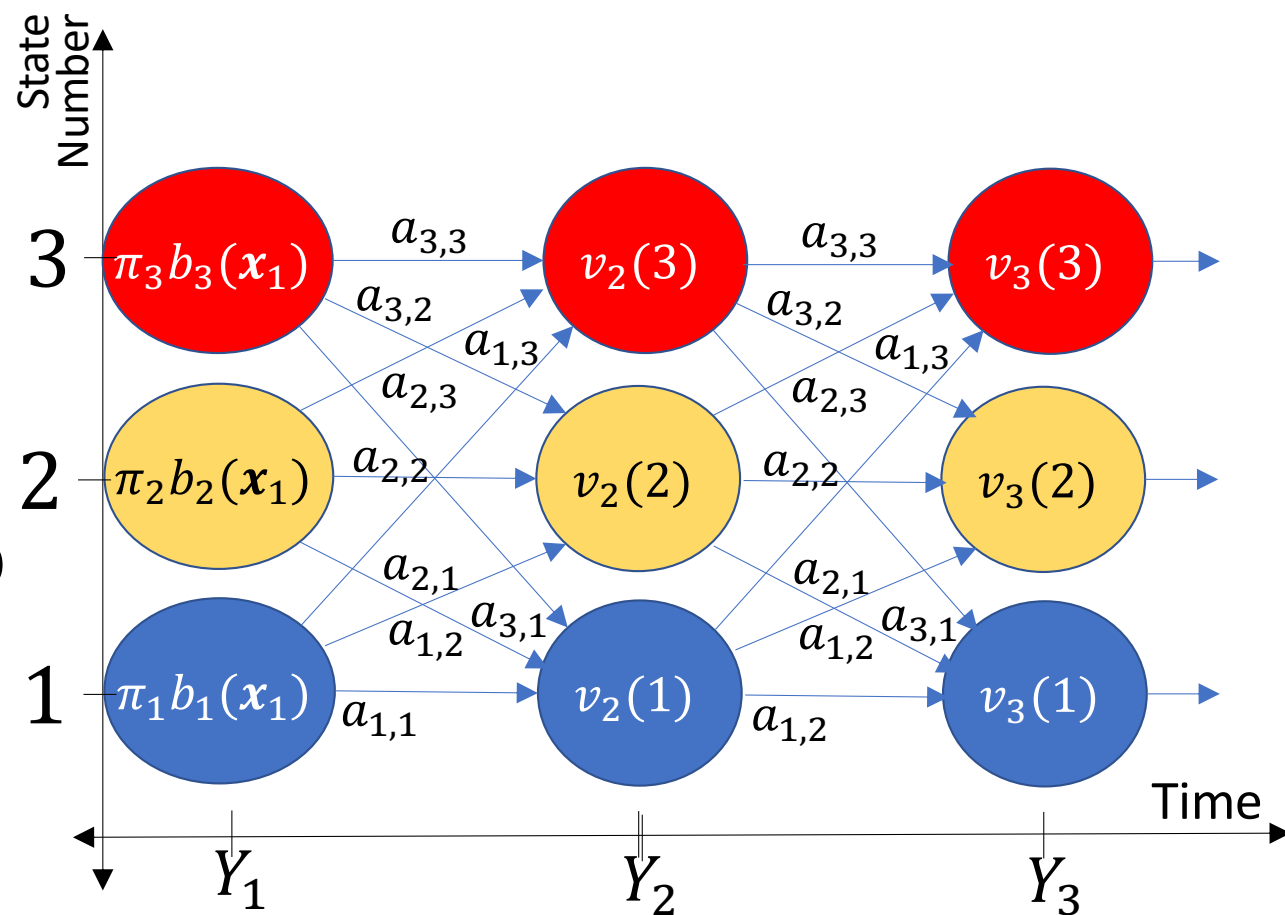
- Positional encoding

$$\mathbf{x}_t += \begin{bmatrix} \cos\left(\frac{\pi t}{T}\right) \\ \sin\left(\frac{\pi t}{T}\right) \\ \vdots \end{bmatrix}$$

# HMM: Viterbi Algorithm

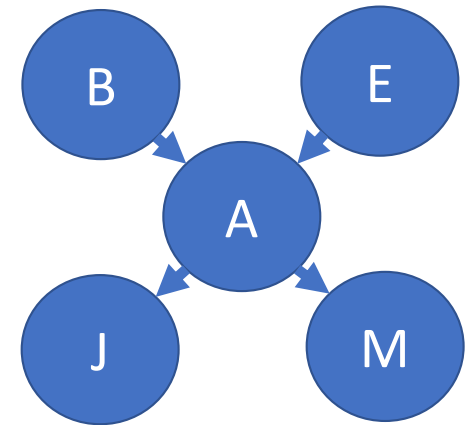
$$v_t(j) = \max_i v_{t-1}(i) a_{i,j} b_j(x_t)$$

$$\psi_t(j) = \operatorname{argmax}_i v_{t-1}(i) a_{i,j} b_j(x_t)$$



# Bayesian Networks

- Bayesian network: A better way to represent knowledge
  - Reduces space complexity from  $\mathcal{O}\{v^n\}$  to  $\mathcal{O}\{nv^p\}$
  - Does not automatically reduce time complexity.
- Key ideas: Independence and Conditional independence



		Shared Ancestor (of at least one)?	
		No	Yes
Shared Descendant (of both)?	No	Independent	Dependent unless shared ancestor value is known
	Yes	Independent unless shared descendant value is known	Dependent unless shared ancestor value known AND shared descendant value unknown

# Explainable AI

- Relevance of input  $x_d$  to output  $z_c$ :

$$R_{c,d} = \frac{\partial z_c}{\partial x_d} \cdot x_d$$

- Layer-wise relevance propagation: Normalize so  $\sum_d R_{c,d} = R_c$
  - Gradient-weighted class activation mapping: Keep only positive relevances
- Counter-Factual Reasoning:

$$\max_{F,X} |P(F, X, A = a) - P(F, X, A = \neg a)| > \epsilon?$$

- If the answer is yes, then you should add an edge between A and F

# 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

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](#)