

430.457
Introduction to Intelligent Systems

Midterm

Prof. Songhwai Oh

Fall 2025

Student Name:

Student ID:

1	/ 20
2	/ 40
3	/ 40
Total	/ 100

IMPORTANT:

- Read all problems carefully and start with easier ones.
- You need to show all your steps for full credit.

Problem 1: Short Answers [20 pts]

- (a) What is overfitting? Describe approaches for reducing the effects of overfitting. [4pt]
- (b) What are two views on the semantics of Bayesian networks? [4pt]
- (c) Show that direct sampling gives a consistent estimate. [4pt]
- (d) Consider a Markov chain $\mathbf{X}_t$ for $t = 0, 1, 2, \dots$. Describe the (first-order) Markov assumption mathematically. [4pt]
- (e) Suppose that there are two events, E_A and E_B , and two observations, O_1 and O_2 . We do not know which observation is from which event. However, we can compute $P(O_i|E_A)$ (or $P(O_i|E_B)$) if we know O_i is an observation about E_A (or E_B). Show how $P(O_1, O_2|E_A, E_B)$ can be computed. [4pt]

(blank page)

Problem 2: [40 pts]

After you aced Introduction to Intelligent Systems, you became a top technology consultant on lithium-ion batteries for the government agencies. After the recent accident in the server room at the National Information Resources Service, you are tasked to design a new fire alarm system for lithium-ion batteries to warn the server room manager about high battery temperature. The alarm should sound if the battery temperature meter reports a high temperature and the battery temperature meter measures the temperature of the battery. For this problem, consider the following Boolean variables: A (alarm sounds), F_A (alarm is faulty), and F_M (meter is faulty) and also consider variables M (battery temperature meter) and B (battery temperature).

- (a) Draw a Bayesian network for this problem.
- (b) Write down the joint distribution of all variables.
- (c) Suppose there are just two possible temperature values for B and M , i.e., normal and high. If the meter is working correctly, it reports the correct temperature with probability a . But if the meter is faulty, it reports the correct value with probability b . Write down the conditional probability table for node M .
- (d) Suppose the alarm works correctly unless it is faulty, in which case it never sounds. Write down the conditional probability table for node A .
- (e) Suppose the alarm and meter are working correctly and the alarm is sounded. Find an expression for the probability that the battery temperature is high, in terms of conditional probabilities in the network.

(blank page)

(blank page)

Problem 3: [40 pts]

Consider a regression problem given N data points $\mathcal{D} = (\mathbf{x}_n, y_n)_{n=1}^N$, where $\mathbf{x}_n \in \mathbb{R}^p$ and $y_n \in \mathbb{R}$, using a hypothesis space in a form of

$$h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x_1 + \cdots + w_p x_p.$$

It is desired to minimize the following l_2 -norm loss function

$$Loss(h_{\mathbf{w}}) = \frac{1}{2} \sum_{n=1}^N (y_n - h_{\mathbf{w}}(\mathbf{x}_n))^2.$$

- (a) Derive the gradient descent update equation for w_i .
- (b) Let $\mathbf{w} = [w_0 \ w_1 \ \cdots \ w_p]^T$. Find $\mathbf{w}^*$ which minimizes the loss function.
- (c) To avoid overfitting, we introduce a regularization term

$$\frac{\lambda}{2} \sum_{i=0}^p w_i^2 = \frac{\lambda}{2} \|\mathbf{w}\|^2$$

added to the loss function, where λ is a user defined constant. Find $\mathbf{w}^*$ which minimizes the regularized loss function.

- (d) Explain the benefits of using the l_1 -norm based regularization term

$$\frac{\lambda}{2} \sum_{i=0}^p |w_i|$$

compared to the regularization term used in (c).

(blank page)

(blank page)