PURDUE

BME 64600 – 001 and ECE 60146 – 001

Midterm #1, March 5, Spring 2026

NAME ______________________________________

PUID ____

Exam instructions:

- You have 75 minutes to work the exam.
- This is a closed-book and closed-note exam. You may not use or have access to your book, notes, any supplementary reference, a calculator, or any communication device including a cell-phone or computer.
- You may not communicate with any person other than the official proctor during the exam.
- There are 32 sub-problems each worth 5pts, for a total score of 160.

To ensure Gradescope can read your exam:

- Write your full name and PUID above and on the top of every page.
- Answer all questions in the area designated for each problem.
- Write only on the front of the exam pages.
- DO NOT run over to the next question.

Name/PUID: Key

Problem 1. (40pt) Probability and Random Variables

Let X, Y, and Z be random variables such that $E[|X|] = E[|Y|] = E[|Z|] < \infty$ on the probability space $(\Omega, \mathcal{B}, P)$ where $\Omega = [0,1]$ and $P(A) = \int_{\omega \in A} d\omega$.
Also, let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, and let $A = [0.5, 0.6]$, $B = [0.1, 0.15]$

a) What is $P(A)$ and $P(B)$?

Solution: $P(A) = 0.1$ and $P(B) = 0.05$.

---

b) Are A and B independent?

Solution: No. A and B are independent if $P(A \cap B) = P(A)\,P(B)$. But $P(A \cap B) = 0 \neq (0.1)(0.05)$, so they are not independent.

---

c) Is $E[X]$ a random variable or number?

Solution: $E[X]$ is a number.

---

d) Is $E[X|Y]$ a random variable or number?

Solution: It is a random variable. In particular, it is a random variable with the form $Z = g(Y)$ for some measurable function g.

---

e) A function $T^*(y)$ has the property that for all (measurable) functions $T : \Re \to \Re$,

$$E\left[(X - T^*(Y))^2\right] \leq E\left[(X - T(Y))^2\right].$$

What exactly is $T^*(Y)$ equal to?
What do we call $T^*(Y)$?

Solution: $T^*(Y) = E[X|Y]$. We call it the minimum mean squared estimate (MMSE).

---

f) Is $f(X)$ a random variable or number?

Solution: Yes, $Z = f(X) = f(X(\omega))$, so it is a function of ω. So Z must be a random variable.

---

g) What is $E[X|X]$ equal to?

Solution: $E[X|X] = X$.

h) What is $E[E[Y|X]]$ equal to?

Solution: $E[E[Y|X]] = E[Y]$

Name/PUID: ______________________________

Problem 2. (30pt) Convexity and Optimization

Consider the functions $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R}^2 \to \mathbb{R}$, and $h : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\begin{aligned} f(x) &= \exp\{-x\} \\ g(x, y) &= x^2 + y^2 \\ h(x, y) &= x^2 - y^2 \ . \end{aligned}$$

a) Is f convex on $\Re$? Justify your answer.

Solution: Yes. It is a continuously differentiable function and $\frac{d^2 f}{dx^2} = e^{-x} > 0$.

---

b) Does f take on a global minimum or local minimum on $\Re$? Justify your answer.

Solution: No, it does not take on a local or global minimum since $\lim_{x \to \infty} f(x) = 0$, but $f(x) > 0$.

---

c) Is g convex on $\Re^2$? Justify your answer.

Solution: Yes, g is convex because it is continuously differentiable and

$$\nabla \nabla g(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

which has eigenvalues of $\lambda_0 = 1$ and $\lambda_1 = 1$, which are both non-negative.

---

d) Does g take on a global minimum, local minimum, or saddle point on $\Re^2$? Justify your answer.

Solution: Yes, g takes on both a local and global minimum for $(x, y) = (0, 0)$ because it is convex and $\nabla g(0) = 0$. It has no saddle points.

---

e) Is h convex on $\Re^2$? Justify your answer.

Solution: No, h is not convex because it is continuously differentiable and

$$\nabla \nabla g(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

which has eigenvalues of $\lambda_0 = 1$ and $\lambda_1 = -1$, which are not both non-negative.

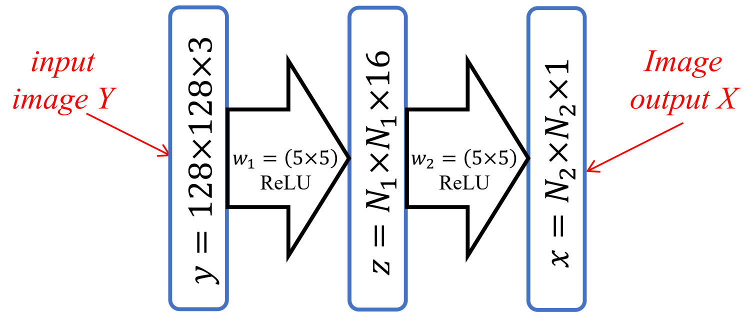
f) Does h take on a global minimum, local minimum, or saddle point on $\Re^2$? Justify your answer.

Solution: No, h does not take on a local or global minimum since $\lim_{y\to\infty} h(0, y) = -\infty$. However, $(x, y) = (0, 0)$ is a saddle point since $\nabla h(0) = 0$.

Name/PUID: ______________________________

Problem 3. (30pt) Convolutional Neural Networks

Consider the following convolutional neural network pictured below with a color image as the input and a gray scale image as the output. Each layer uses a ReLu activation function, a **"valid"** boundary condition, and denote the convolution kernels by w_1 and w_2 and the associated offsets by b_1 and b_2.



a) Calculate the value of N_1.

Solution: $N_1 = 124 = 128 - (5 - 1)$

---

b) Calculate the value of N_2.

Solution: $N_2 = 120 = 124 - (5 - 1)$

---

c) Calculate the shape of w_1.

Solution: $5 \times 5 \times 3 \times 16$

---

d) Calculate the shape of w_2.

Solution: $5 \times 5 \times 16 \times 1$

---

e) Calculate the shape of b_1.

Solution: 16

---

f) Calculate the total number of parameters in the model.
(Hint: Show all your work so if you multiple incorrectly, you'll still get most of the points.)

Solution: Number of parameters in each layer:

- layer1:
  - filter: $5 \times 5 \times 3 \times 16 = 1200$
  - offset: 16
- layer2:
  - filter: $5 \times 5 \times 16 \times 1 = 400$
  - offset: 1

Total number of parameters:

$$1200 + 16 + 400 + 1 = 1617$$

Name/PUID: ______________________________

Problem 4. (25pt) CNN Adjoint Gradient

Let $x = f(y, w)$ represent a 1D convolution block using "valid" boundary conditions with input $y = [y_0, \cdots, y_{N-1}]$ and convolution kernel $w = [w_0, w_1, w_2]$.
Furthermore, define:

$f_w(y)$ is the linear operator taking y to x;

$f_y(w)$ is the linear operator taking w to x;

f_w^* is the adjoint of f_w;

f_y^* is the adjoint of f_y.

Note: For this exam, assume that this is **true convolution** rather than the correlation typically used in most software.

a) If $f_w(y) = A_w y$ where A_w denotes a matrix, then what is the shape of A_w?

Solution: A_w is an $(N-2) \times N$ matix.

---

b) If $f_w(y) = A_w y$ where A_w denotes a matrix, then write out the form of the matrix A_w for arbitrary N.

Solution:

$$A_w = \begin{bmatrix} w_2 & w_1 & w_0 & 0 & \cdots & 0 \\ 0 & w_2 & w_1 & w_0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & w_2 & w_1 & w_0 & 0 \\ 0 & \cdots & 0 & w_2 & w_1 & w_0 \end{bmatrix}$$

---

c) If $f_w^*(z) = B_w z$ where B_w denotes a matrix, then what is the shape of the matrix B_w?

Solution: B_w is a matrix with shape $N \times (N-2)$.

---

d) If $f_w^*(z) = B_w z$ where B_w denotes a matrix, then write out an explicit expression for B_w.

Solution:

$$B_w = \begin{bmatrix} w_2 & 0 & 0 & \cdots & 0 \\ w_1 & w_2 & 0 & \cdots & 0 \\ w_0 & w_1 & & & \vdots \\ 0 & w_0 & & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & w_2 & 0 \\ 0 & 0 & & w_1 & w_2 \\ 0 & 0 & & w_0 & w_1 \\ 0 & 0 & & 0 & w_0 \end{bmatrix}$$

---

e) If $f_y(w) = C_y w$ where C_y denotes a matrix, then what is the shape of C_y?

Solution: C_y is an $(N-2) \times 3$ matrix.

---

f) If $f_y(w) = C_y w$ where C_y denotes a matrix, then write out the form of the matrix C_y for arbitrary N.

Solution:

$$A_w = \begin{bmatrix} y_2 & y_1 & y_0 \\ y_3 & y_2 & y_1 \\ \vdots & \vdots & \vdots \\ y_{N-1} & y_{N-2} & y_{N-3} \end{bmatrix}$$

---

g) If $f_y^*(z) = D_y z$, then what is the shape of the matrix D_y?

Solution: D_y is a matrix with shape $3 \times (N-2)$.

---

h) If $f_y^*(z) = D_y z$, then write out the form of the matrix D_y for arbitrary N.

Solution:

$$D_y = \begin{bmatrix} y_2 & y_3 & y_4 & \cdots & y_{N-1} \\ y_1 & y_2 & y_3 & \cdots & y_{N-2} \\ y_0 & y_1 & y_2 & \cdots & y_{N-3} \end{bmatrix}$$

---

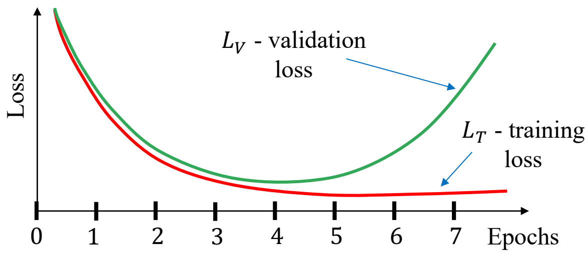
i) What is the interpretation of the function $f_y^*(z) = D_y z$?

Solution: This is correlation of the input signal $z = [z_0, \cdots, z_{N-3}]$ with the signal $y = [y_0, \cdots, y_{N-1}$ using a "valid" boundary condition. The "valid" boundary condition ensures that there are only three output points corresponding to the three weights.

Name/PUID: ______________________________

Problem 5. (25pt) Training, generalization, and regularization

You are training a deep neural network, and you get the result pictured below.



a) What number of epochs is likely to the result in the best parameter estimates? Why?

Solution: 4 epochs because this is the value at which the validation loss is approximately minimized.

---

b) What is likely to happen if you choose the parameter estimates from 7 epochs? Why?

Solution: In this case, the parameter estimates will likely be overfit to the training data, so they will not generalize well.

---

c) What does this curve suggest about the capacity of your model?

Solution: The curve suggests that the model's capacity may be a bit too high given the amount of available training data.

---

d) Given your answer to c) above and assuming that you can not change the model, then what other modification to the training might be helpful? Why?

Solution: It might be helpful to introduce regularization of the parameters into the loss function. For example, l_1 regularization can enforce sparsity in the model, there by removing the unnecessary parameters, and effectively reducing the model order.

---

e) If you increase the training data by a factor of 10, then what would you expect to happen to the **validation** loss curve?

Solution: Since the larger amount of training data will reduce overfitting, I would expect the validation loss to go down.
