CSS324: Artificial Intelligence

Final Mock Exam

curated by The Peanuts

Name.............ID.............Section.............Seat No.............

Conditions: Semi-closed Book

Directions:

1. This exam contains 14 pages (including this one).
2. Write your name clearly at the top.
3. This is a semi-closed book exam. "Semi" means you may open your book halfway, but if you open it fully, we will apply dropout (to you).
4. If your brain crashes, please reboot by drinking water. Memory leaks are common during finals season.
5. Marking scheme uses softmax. The examiner cannot guarantee numerical stability.
6. If you take too long on one question, early stopping will be applied without warning.

For solution, ***click here****.*

Problem 1

For each statement below, write **TRUE** or **FALSE** in the space provided.

1.1 _______________ A feedforward neural network with a single hidden layer can represent any function, though it may require an infeasibly large number of units.

1.2 _______________ He initialization is specifically designed for networks using sigmoid or tanh activation functions.

1.3 _______________ In a CNN, max pooling layers have trainable parameters that are updated during backpropagation.

1.4 _______________ The hidden state at time step t in an RNN depends on both the current input and the previous hidden state.

1.5 _______________ L1 regularization is more likely than L2 regularization to produce sparse weight matrices with many weights equal to zero.

1.6 _______________ In an LSTM, the cell state acts as a long-term memory that can carry information across many time steps.

1.7 _______________ Cross-entropy loss is typically used for regression tasks where the output is a continuous value.

1.8 _______________ In a Transformer model, the self-attention mechanism allows the model to process all positions in parallel, unlike RNNs which must process sequentially.

1.9 _______________ Global average pooling reduces each feature map to a single value by computing the average across all spatial locations.

1.10 _______________ In a GAN, the generator is trained to minimize the probability that its generated output is classified as real by the discriminator.

Problem 2

A detective named Sherlock is investigating a mysterious disappearance at a Bangkok mansion. He knows the following:

- The mansion has a security system that may be activated or not.
- A wealthy tycoon lives there who may be home or away.
- The guard will call the police if suspicious.
- A maid also calls police if something seems wrong.
- The security system activation depends on whether the tycoon is home.
- Both the guard and maid's behavior are influenced by whether the security system is triggered.
- The guard's behavior is also influenced by whether the tycoon is home.

Given the following variables:

- T = Tycoon is home
- S = Security system is on
- G = Guard calls police
- M = Maid calls police

The probabilities are:

$$p(t) = 0.7$$

$$\begin{aligned} p(s|t) &= 0.3, \quad p(s|\neg t) = 0.8 \\ p(g|t,s) &= 0.95, \quad p(g|t,\neg s) = 0.05 \\ p(g|\neg t,s) &= 0.98, \quad p(g|\neg t,\neg s) = 0.15 \\ p(m|s) &= 0.85, \quad p(m|\neg s) = 0.10 \end{aligned}$$

Answer the following questions based on this model:

2.1) Construct a Bayesian network that represents the dependencies among these variables.

2.2) Write the joint probability distribution for all four variables using the chain rule based on the Bayesian Network structure.

2.3) What is the probability that the security system is on, the guard calls police, the tycoon is away, and the maid does not call police?

2.4) Given that the guard called the police, what is the probability that the security system was on?

Problem 3

3.1) You are building a neural network for a regression task to predict house prices. What should you use for the output layer?

- Number of neurons: _______________
- Activation function: _______________
- Loss function: _______________

3.2) You have a dataset with 5 categorical values for a "weather" feature: {sunny, cloudy, rainy, snowy, foggy}. How would you encode this for input to a neural network? Describe your encoding scheme and show an example for "rainy".

3.3) You are training a deep neural network and observe the following:

- Training loss: 0.05 (very low)
- Validation loss: 0.45 (high)
- Test loss: 0.42 (high)

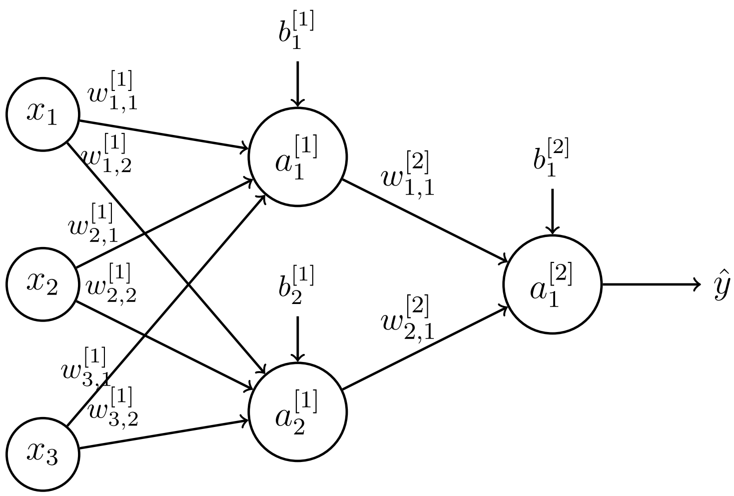
(a) What problem is your model experiencing?

(b) Suggest TWO techniques you could use to address this problem.

3.4) Explain the difference between He initialization and Xavier initialization. When would you use each one?

Problem 4

Consider a multilayer perceptron with the following structure:



All artificial neurons use sigmoid activation functions. The network parameters are:

$$\mathbf{W}^{[1]} = \begin{bmatrix} 0.5 & -0.2 & 0.3 \\ 0.3 & 0.4 & -0.1 \end{bmatrix}, \quad \mathbf{b}^{[1]} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$$

$$\mathbf{W}^{[2]} = \begin{bmatrix} 0.8 & -0.6 \end{bmatrix}, \quad b^{[2]} = 0.3$$

Given a training example:

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad y = 1$$

The binary cross-entropy loss function is:

$$L(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid activation function.

4.1) Perform the forward pass through the network. Calculate and show:

(i) $z_1^{[1]}, z_2^{[1]}$ (the weighted sums for hidden layer neurons)

(ii) $a_1^{[1]}, a_2^{[1]}$ (the activations of hidden layer neurons)

(iii) $z^{[2]}$ (the weighted sum for output neuron)

(iv) $\hat{y} = a_1^{[2]}$ (the final output)

4.2) Calculate the loss $L(y, \hat{y})$ for this training example. Show your work.

4.3) Write the chain rule expression for $\frac{\partial L}{\partial w_{1,1}^{[1]}}$. Express your answer as a product of partial derivatives. You do NOT need to compute the numerical value.

4.4) Suppose you decide to use the Adam optimizer with learning rate $\eta = 0.001$ instead of vanilla SGD. Briefly explain what are the two main components that Adam uses to improve convergence, and how they differ from standard gradient descent.

Problem 5

5.1) Calculate the output shape after applying a convolutional layer with the following specifications:

- Input shape: $32 \times 32 \times 3$
- Number of filters: 64
- Kernel size: 5×5
- Stride: 1
- Padding: 2 (same padding)

Show your calculation and final answer.

5.2) After the convolution in part (a), a max pooling layer is applied with:

- Pool size: 2×2
- Stride: 2

What is the output shape after this pooling layer?

5.3) **(5 points)** Calculate the number of parameters (weights and biases) in the convolutional layer from part (a).

5.4) **(4 points)** Consider the following PyTorch code for a simple CNN:

```
class SimpleCNN(nn.Module):
    def __init__(self):
        super(SimpleCNN, self).__init__()
        self.conv1 = nn.Conv2d(1, 6, 5, padding=2)
        self.conv2 = nn.Conv2d(6, 16, 5, padding=0)
        self.flatten = nn.Flatten()
        self.fc1 = nn.Linear(16 * 5 * 5, 120)
        self.fc2 = nn.Linear(120, 84)
        self.fc3 = nn.Linear(84, 10)

    def forward(self, x):  # x: 28×28×1
        x = self.conv1(x)    # 28×28×6
        x = F.avg_pool2d(x, 2)  # 14×14×6
        x = self.conv2(x)    # ?×?×16
        x = F.avg_pool2d(x, 2)  # ?×?×16
        x = self.flatten(x)
        x = self.fc1(x)
        x = F.relu(x)
        x = self.fc2(x)
        x = F.relu(x)
        x = self.fc3(x)
        return x
```

Fill in the missing dimensions marked with "?" after `self.conv2(x)` and the second `avg_pool2d`.

Problem 6

A cryptocurrency trader wants to predict Bitcoin price movements using a simple RNN. The network tracks daily price changes (in thousands of USD) to forecast the next day's movement.

RNN Parameters:

$$\begin{aligned} w_x &= 0.4 \\ w_h &= 0.25 \\ w_y &= 2.0 \\ b_h &= -0.5 \\ b_y &= 1.0 \\ h_0 &= 0 \end{aligned}$$

Model equations:

$$\begin{aligned} h_t &= w_x x_t + w_h h_{t-1} + b_h \quad \text{(identity activation)} \\ \hat{y}_t &= w_y h_t + b_y \end{aligned}$$

where x_t represents the price change on day t, and $\hat{y}_t$ is the predicted price change for the next day.

6.1) Over three consecutive days, Bitcoin experienced the following price changes:

- Day 1: $x_1 = +3.0$ (price increased by \$3,000)
- Day 2: $x_2 = -1.5$ (price decreased by \$1,500)
- Day 3: $x_3 = +2.0$ (price increased by \$2,000)

Calculate:

(i) h_1, h_2, h_3 (show your work for each step)

(ii) The predicted price change for Day 4

6.2) The trader wants to extend the prediction window to forecast prices 30 days ahead. Explain why vanilla RNNs would struggle with this task.

6.3) The trader decides to use an LSTM instead. Briefly explain how the LSTM's gating mechanisms (forget gate, input gate, and cell state) help maintain long-term memory for better long-range predictions.