Name:

NetID:

1.) What are some of the benefits of self-attention when compared to RNNs? (1 pt.)

2.) What is a filter used for in a Convolutional Neural Network? (1 pt.)

3.) What are over and underfitting? (2 pts.)

4.) Given the pytorch code below, answer the following questions.

```
1  import torch
2  import torch.nn as nn
3  class Model4_1(nn.Module):
4      def __init__(self):
5          super(Model4_1, self).__init__()
6          self.lin1 = nn.Linear(784, 100)
7          self.relu = nn.ReLU()
8          self.lin2 = nn.Linear(100, 10)
9          
10     def forward(self, x):
11         out = self.lin1(x)
12         out = self.relu(out)
13         out = self.lin2(out)
14         return out
```

Note: You can assume softmax and cross entropy loss are used after the final layer, the two yielding a combined derivative of: ŷ - y

a.) Draw the computation graph for the network. (4 pts.)

b.) Give the derivative chain for calculating the gradient of the bias in the first layer. (3 pts.)

$$\frac{\partial \mathcal{L}}{\partial b} =$$

c.) Give the equation you'd use to actually calculate the gradients. (2 pts.)

d.) Given the gradients, what's the next step we should take? (1 pt.)

e.) What task could this network be used for? (Hint: It's something I used as an example a lot) (1 pt.)

5.) Why do we want activation functions in neural networks? (1 pt.)

6.) How could you use a kNN method for a regression task? (1 pt.)

7.) What's an example of a feature which would cause us to want to use a non-Gaussian distribution in a Naive Bayes classifier? (1 pt.)

8.) What is a feature vector, why are these useful? (2 pts.)

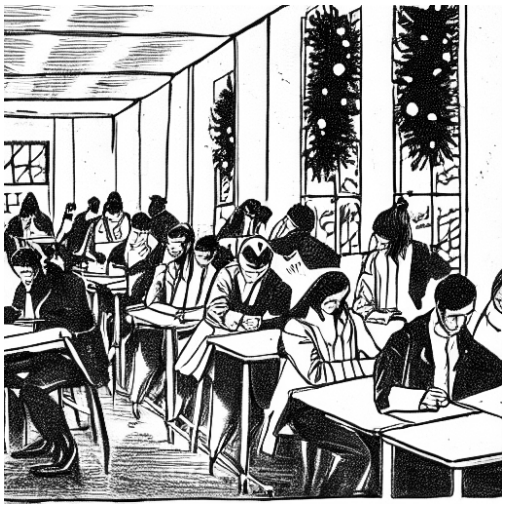
9.) I have a bunch of anonymized student exam grades, which model should I use to build a classifier that would allow me to predict which student got which grade? (2 pts.)

10.) A new tech company in Norway is training their faceID tech on participants in yoga classes. They're about to roll it out globally, do you foresee any issues with this? (2 pts.)

11.) When we say transformers are autoregressive what are we saying? (1 pt.)

12.) When we get a sentence for an NLP task there are two steps before we can pass it into our model, what are these and what do they “look” like? (2 pts.)

Breather.) Take a breath and rest a little. Enjoy this image of “black and white line drawing of students acing their final exam beneath a christmas tree” I generated using the simple stable diffusion model we looked at in class. Halfway there, I believe in you, you got this! (2 pts.)



13.) Given the training code for a pytorch model below, answer the following questions:

```
# Instantiate the model, define loss and optimizer
model = SimpleNet()
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=1.00)

# Training loop
num_epochs = 1
train_losses = []

for epoch in range(num_epochs):
    model.train()
    optimizer.zero_grad()

    # Forward pass
    outputs = model(X_train)
    loss = criterion(outputs, y_train)

    # Backward pass and optimization
    loss.backward()
    optimizer.step()
```

a.) My model is really bad, what are at least two things I could change that may make it better? (2 pts.)

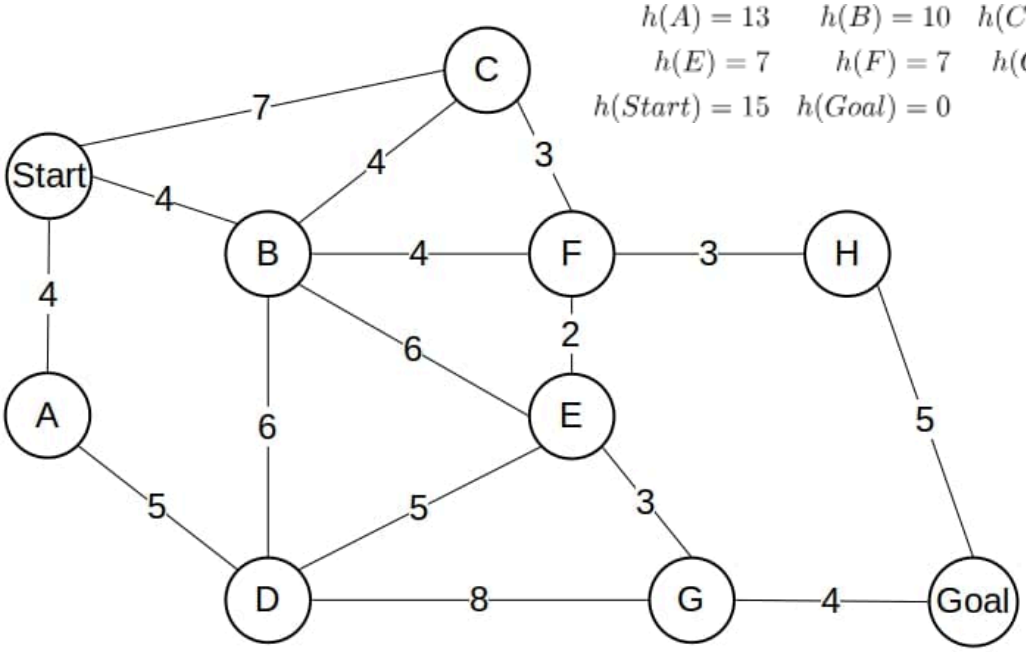
b.) What are the names for what the lines loss.backward() and optimizer.step do? (2 pts.)

14.) On Homework 0910, what purpose did <PAD> serve? (2 pts.)

15.) On Homework 0910, Why did we pass the output of an LSTM into a fully connected layer? (2 pts.)

16.) Given the following graph, starting in Node S and trying to get to Node G, trace the steps the A* algorithm would take. To make the graph less busy, the heuristic distances are given as a table. Much like on Exam 01, my solution uses all 11 spots. (8 pts.)

$h(A) = 13$ $h(B) = 10$ $h(C) = 10$ $h(D) = 9$
$h(E) = 7$ $h(F) = 7$ $h(G) = 2$ $h(H) = 3$
$h(Start) = 15$ $h(Goal) = 0$



*.) N -> N' = d(N, N') + h(N', G)

1.) S -> A = 4 + 13 = 17

2.) S -> B = 4 + 10 = 14

3.) S -> C = 7 + 10 = 17

4.) S ->

5.) S ->

6.) S ->

7.) S ->

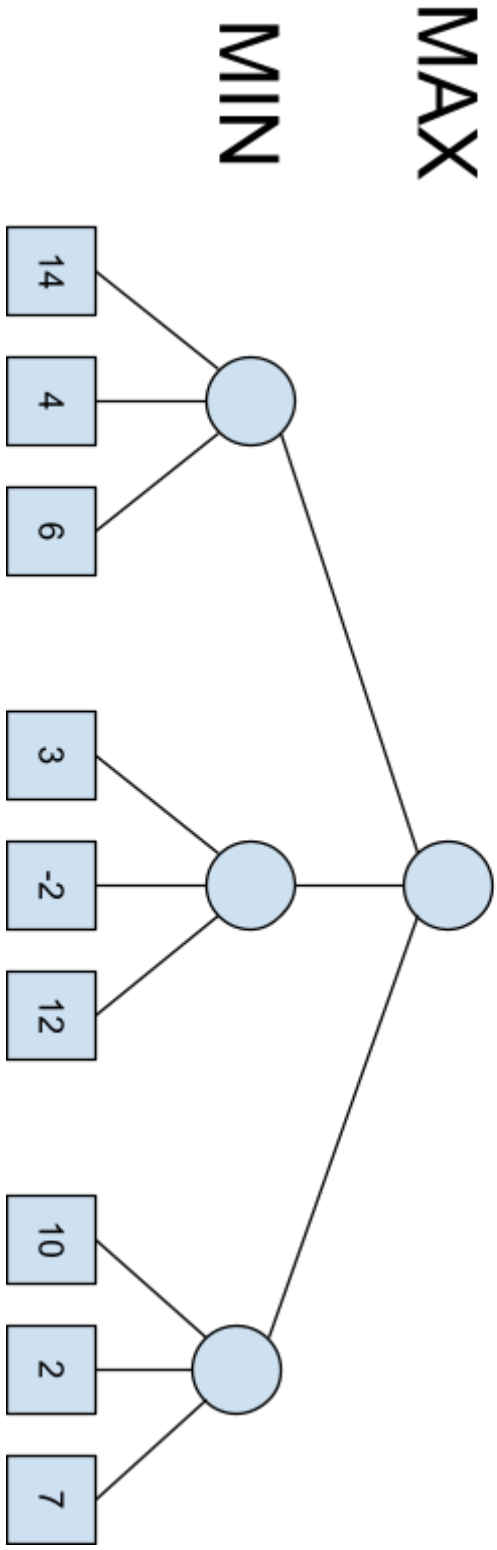
8.) S ->

9.) S ->

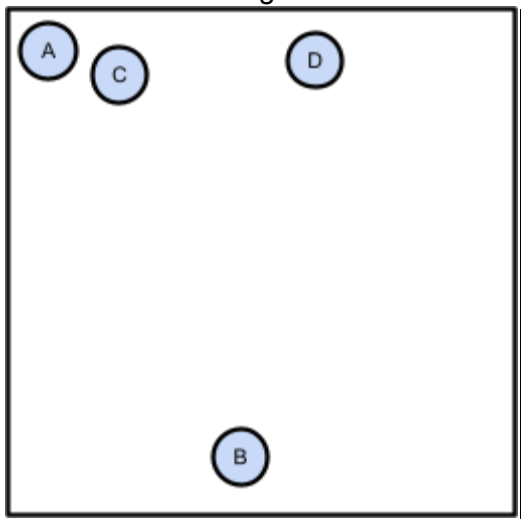
10.) S ->

11.) S ->

17.) Use Alpha/Beta Pruning on the tree below, showing the Alpha and Beta at each node and indicate which branches don't need to be explored. Make sure to circle the final score. (12 pts.)



18.) Given the data points, draw the dendrogram that would be created using agglomerative hierarchical clustering and then draw a line on the dendrogram to create 2 clusters. (3 pts.)



19.) Given the trivariate Jacobian matrix below you can compute the gradients for cost function J. Given the values 1, 3, 2 for the 3 variables nah I'm just kidding have a Merry Christmas and if you write down or draw what you're most excited to do over break I'll give 2 bonus points. Thanks for the semester! Ignore the next sentence. Solve the Jacobian for the gradients. (2 pts.)

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

$$= \frac{\partial x}{\partial u}\begin{vmatrix} \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix} - \frac{\partial x}{\partial v}\begin{vmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial w} \end{vmatrix} + \frac{\partial x}{\partial w}\begin{vmatrix} \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{vmatrix}$$

$$= \frac{\partial x}{\partial u}\left(\frac{\partial y}{\partial v}\cdot\frac{\partial z}{\partial w} - \frac{\partial y}{\partial w}\cdot\frac{\partial z}{\partial v}\right) - \frac{\partial x}{\partial v}\left(\frac{\partial y}{\partial u}\cdot\frac{\partial z}{\partial w} - \frac{\partial y}{\partial w}\cdot\frac{\partial z}{\partial u}\right) + \frac{\partial x}{\partial w}\left(\frac{\partial y}{\partial u}\cdot\frac{\partial z}{\partial v} - \frac{\partial y}{\partial v}\cdot\frac{\partial z}{\partial u}\right)$$