Quiz 4

Graded

Student

Alaa Alghwiri

Total Points

5 / 8 pts

Question 1

Linear Neural Networks

1 / 1 pt

✔ **+ 1 pt** Correct

+ 0 pts Incorrect

Question 2

Output Units and loss functions

0 / 1 pt

+ 1 pt Correct

✔ **+ 0 pts** Incorrect

Question 3

Regression Output Units

1 / 1 pt

3.1 **Real Values**

0.25 / 0.25 pts

✔ **+ 0.25 pts** Correct

+ 0 pts Incorrect

3.2 **regression [0,1]**

0.25 / 0.25 pts

✔ **+ 0.25 pts** Correct

+ 0 pts Incorrect

3.3 **regression [-1,1]**

0.25 / 0.25 pts

✔ **+ 0.25 pts** Correct

+ 0 pts Incorrect

3.4 **[-10, 5]**

0.25 / 0.25 pts

✔ **+ 0.25 pts** Correct

+ 0 pts Incorrect

Question 4

Bounded output units

0 / 2 pts

+ 2 pts Correct

✔ + 0 pts Incorrect

Question 5

Activation Functions

1 / 1 pt

✔ + 1 pt Correct

+ 0 pts Incorrect

Question 6

Softmax

1 / 1 pt

✔ + 1 pt Correct

+ 0 pts Incorrect

Question 7

Depth versus Width

1 / 1 pt

✔ + 1 pt Correct

+ 0 pts Incorrect

Q1 Linear Neural Networks
1 Point

Consider an input space of $x \in \mathbb{R}^3$ and a linear neural network with 3 hidden layers each with 16 hidden units. If we increase the number of layers (depth) of the network or increase the width of each layer then this neural network will be able to represent more functions

- ○ true
- ◉ false

Q2 Output Units and loss functions
1 Point

Consider the different combinations of output units and the loss functions. Which combinations will lead to better predictions when minimizing the loss function.

- ☐ Classification: mean squared error and sigmoid units
- ☐ Classification: negative probability and sigmoid units
- ☑ Classification: negative log probability and sigmoid units
- ☐ Classification: mean squared error and linear units
- ☐ Classification: negative probability and linear units
- ☐ Classification: negative log probability and linear units
- ☐ Regression: mean squared error and sigmoid units
- ☑ Regression: mean squared error and linear units
- ☐ Regression: negative log density (continuous version of probability) and linear units

Q3 Regression Output Units

1 Point

For regression, we are usually concerned with predicting a scalar $y \in \mathbb{R}$. Regression does not have to be in the space of all real numbers but can be on any bounded subset. Choose the most appropriate output units for each target.

Q3.1 Real Values

0.25 Points

Let targets $y \in \mathbb{R}$. The best output unit is:

- (●) linear
- ( ) sigmoid
- ( ) tanh
- ( ) ReLU

Q3.2 regression [0,1]

0.25 Points

Let targets $y \in [0, 1]$. The best output unit is:

- ( ) linear
- (●) sigmoid
- ( ) tanh
- ( ) ReLU

Q3.3 regression [-1,1]

0.25 Points

Let targets $y \in [-1, 1]$. The best output unit is:

- ( ) linear
- ( ) sigmoid
- (●) tanh
- ( ) ReLU

Q3.4 [-10, 5]
0.25 Points

Let the targets be in $y \in [-10, 5]$. The best output unit is:

- (●) linear
- ( ) sigmoid
- ( ) tanh
- ( ) ReLU

Q4 Bounded output units

2 Points

Consider having the target in some bounded range $y \in (a, b)$, e.g., $(-10, 5)$. If we use a linear output unit, then the network will likely predict values outside this range. Give a brief suggestion of how we could make use of sigmoid or tanh output units.

range of sigmoid and tanh are or might different from [a,b]. We can probably rescale the outcome to match with the required range.

Q5 Activation Functions
1 Point

Recall the properties of vanishing and exploding gradients. Which activation functions below do cause vanishing or exploding gradients?

- [ ] $\sigma(x) = \frac{1}{1+e^{-x}}$
- [ ] $\sigma(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
- [x] $\sigma(x) = \max(0, x)$
- [ ] $\sigma(x) = \ln(1 + e^{x})$
- [x] $\sigma(x) = |x|$
- [x] $\sigma(x) = x$

Q6 Softmax
1 Point

Consider using a softmax to represent a probability distribution over class labels. Mark all statements below that are true

- [ ] can only be used for three or more classes
- [x] can add or subtract a constant to the logits z and not change the probabilities
- [ ] can multiply the logits by a positive constant and not change the probabilities
- [x] can be used with negative log-likelihood loss function

Q7 Depth versus Width
1 Point

Which is more likely to lead to better performance.

- ○ increasing the width of each layer by n units
- ◉ adding another layer with n units
- ○ neither