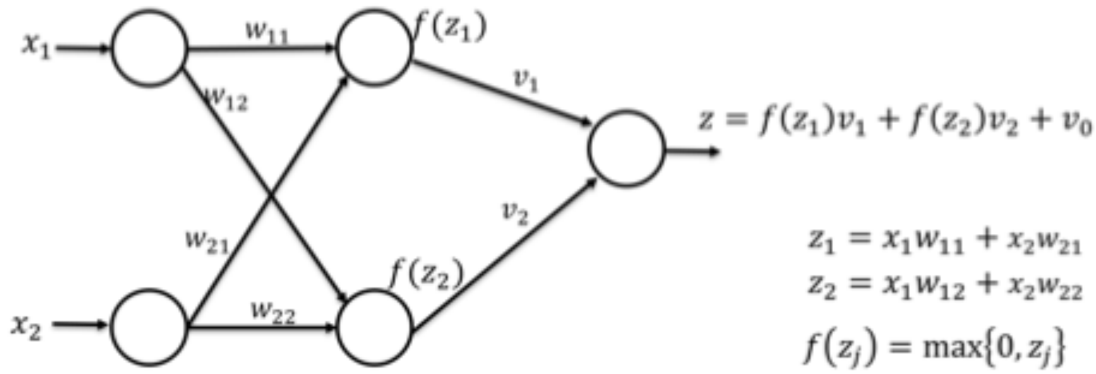
End My Exam 47:54:59

You are taking "Midterm Exam 1" as a timed exam. The timer on the right shows the time remaining in the exam. To receive credit for problems, you must select "Submit" for each problem before you select "End My Exam". **Show Less**

Problem 5

Midterm due Nov 9, 2020 18:59 EST

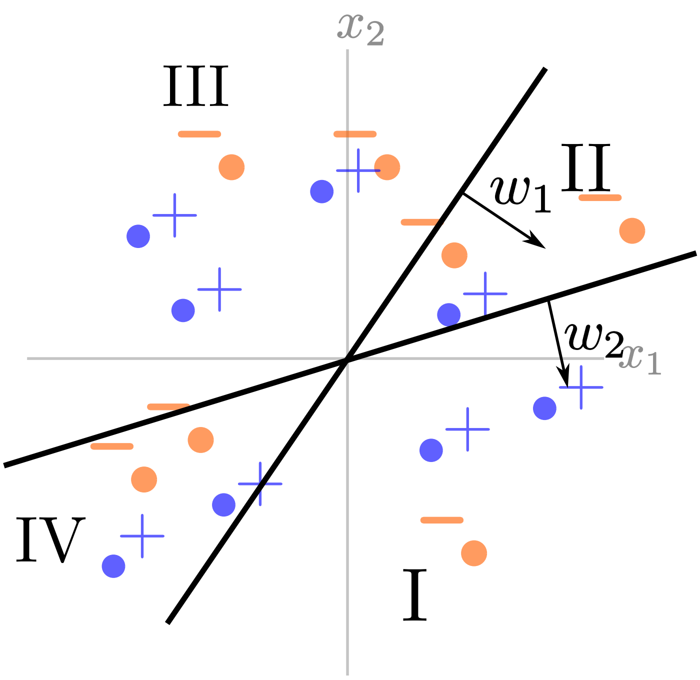
Consider a 2-layer feed-forward neural network that takes in $x \in \mathbb{R}^2$ and has two ReLU hidden units as defined in the figure below. **Note that hidden units have no offset parameters in this problem.**



5. (1)

4 points possible (graded, results hidden)

The values of the weights in the hidden layer are set such that they result in the z_1 and z_2 “classifiers" as shown in the (x_1, x_2)**-space** in the figure below:



The z_1 “classifier" with the normal $w_1 = \begin{bmatrix} w_{11} & w_{21} \end{bmatrix}^T$ is the line given by $z_1 = x \cdot w_1 = 0.$

Similarly, the z_2 “classifier" with the normal $w_2 = \begin{bmatrix} w_{12} & w_{22} \end{bmatrix}^T$ is the line given by $z_2 = x \cdot w_2 = 0.$

The arrows labeled w_1 and w_2 point in the **positive** directions of the respective normal vectors.

The regions labeled I, II, III, IV are the 4 regions defined by these two lines not including the boundaries.

Choose the region(s) in (x_1, x_2) space which are mapped into each of the following regions in (f_1, f_2)-space, the 2-dimensional space of hidden unit activations $f(z_1)$ and $f(z_2)$. (For example, for the second column below, choose the region(s) in (x_1, x_2) space which are mapped into the f_1-axis in (f_1, f_2)-space.)

(Choose all that apply for each column.)

$\{(f_1, f_2) : f_1 > 0,\ f_2 > 0\}$:	f_1-axis:	f_2-axis:	the origin $(f_1, f_2) = (0, 0)$:

(Choose all that apply.)

- [ ] I
- [ ] II
- [ ] III
- [ ] IV
- [ ] None of the above

- [ ] I
- [ ] II
- [ ] III
- [ ] IV
- [ ] None of the above

- [ ] I
- [ ] II
- [ ] III
- [ ] IV
- [ ] None of the above

- [ ] I
- [ ] II
- [ ] III
- [ ] IV
- [ ] None of the above

Submit You have used 0 of 3 attempts

---

5. (2)

2 points possible (graded, results hidden)

If we keep the hidden layer parameters above fixed but add and train additional hidden layers (applied after this layer) to further transform the data, could the resulting neural network solve this classification problem?

- yes
- no

Suppose we stick to the 2-layer architecture but add many more ReLU hidden units, all of them without offset parameters. Would it be possible to train such a model to perfectly separate these points?

Note : Assume that no 2 data points lie on the same line through the origin.

- yes
- no

Submit You have used 0 of 3 attempts

---

5. (3)

5 points possible (graded, results hidden)

Which of the following statements is correct?

1. The gradient calculated in the backpropagation algorithm consists of the partial derivatives of the loss function with respect to each network weight.

   - True
   - False

2. Initialization of the parameters is often important when training large feed-forward neural networks.

   If weights in a neural network with sigmoid units are initialized all the weights to close to zero values, then during early stochastic gradient descent steps, the network represents a nearly linear function of the inputs.

   - True
   - False

3. On the other hand, if we randomly set all the weights to very large values, or don't scale them properly with the number of units in the layer below, then the sigmoid units would behave like sign units. Here, “behave like sign units" allows

for shifting or rescaling of the sign function.

(Note that a sign unit is a unit with activation function $\text{sign}\,(x) = 1$ if $x > 0$ and $\text{sign}\,(x) = -1$ if $x < 0$. For the purpose of this question, it does not matter what $\text{sign}\,(0)$ is.)

- True

Grading Note: (November 1) Since there is an error in this question, i.e. there is only 1 option in the multichoice, everyone will receive credit.

4. If we use only sign units in a feedforward neural network, then the stochastic gradient descent update will

- almost never change any of the weights
- change the weights by large amounts at random

5. Stochastic gradient descent differs from (true) gradient descent by updating only one network weight during each gradient descent step.

- True
- False

Submit You have used 0 of 3 attempts

5. (4)

3 points possible (graded, results hidden)

There are many good reasons to use convolutional layers in CNNs as opposed to replacing them with fully connected layers. Please check T or F for each statement.

Since we apply the same convolutional filter throughout the image, we can learn to recognize the same feature wherever it appears.

- True
- False

A fully connected layer for a reasonably sized image would simply have too many parameters

- True
- False

A fully connected layer can learn to recognize features anywhere in the image even if the features appeared preferentially in one location during training

- True
- False

Submit You have used 0 of 3 attempts

Error and Bug Reports/Technical Issues

Hide Discussion

Topic: Midterm Exam (1 week):Midterm Exam 1 / Problem 5

Add a Post

Show all posts by recent activity

5. (4) "A fully connected layer for a reasonably sized image..."
May I ask to clarify: "reasonably" sized means 1024x1024 or 28x28? Is the reason for sizing was n...
2

Grammatical error - confusing wording
Q5.3.2 - "If weights in a neural network with sigmoid units are initialized all the weights to close to...
1

Question 5. (3) 3.
does not have 2 options, true/false. Is this normal?
2

[STAFF] 5.3.3 does not have 'False' option.
Can you please fix?
4