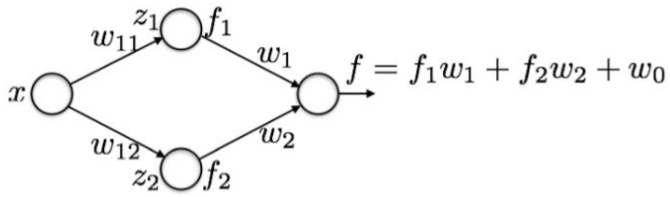
PROBLEM 12

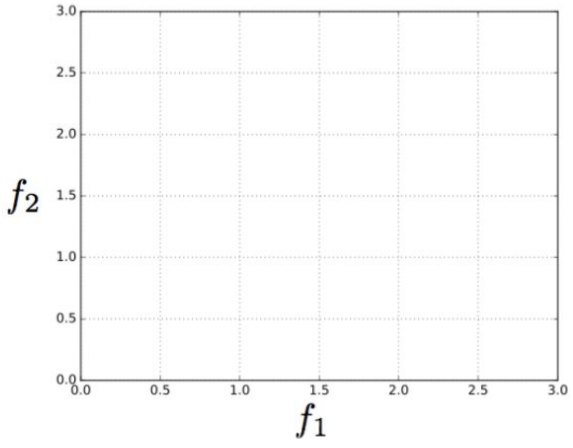
Problem 5 Consider a simple feed-forward neural network model that takes $x \in \mathbb{R}$ as an input and has two ReLU hidden units.



$$\begin{aligned} z_1 &= x\, w_{11} + w_{01} \\ z_2 &= x\, w_{12} + w_{02} \end{aligned}, \quad \begin{bmatrix} w_{11} & w_{01} \\ w_{12} & w_{02} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \tag{14}$$

(5.1) **(2 points)** When is the output of the first hidden unit, i.e., f_1, exactly zero as a function of $x \in \mathbb{R}$?

(5.2) **(6 points)** Given the parameters in Eq(14), sketch in the figure below how examples $x \in [-2, 2]$ map to the 2-dimensional feature coordinates (f_1, f_2). In other words, map the interval $[-2, 2]$ in the x-space to the 2-dimensional space of hidden unit activations.



(5.3) **(2 points)** Are the training examples $(x = -1, y = -1)$, $(x = 1, y = 1)$, and $(x = 2, y = -1)$ linearly separable in the (f_1, f_2) coordinates? (Y/N) **( )**

(5.4) **(4 points)** Suppose $w_1 = 1$, $w_2 = 1$, and $w_0 = 0$. We use $\text{Loss}_h(yf) = \max\{0, 1 - yf\}$ to measure the loss given the network output f and true label y. The parameters will potentially change if we perform a gradient descent step in response to (x, y) where $y = -1$. Provide the range of values of x such that w_{02} *decreases* after the update: $x \in$ (

SKIP

)

(5.5) **(3 points)** The architecture and training details are sometimes quite relevant to how well a particular neural network model works. Suppose we use ReLU activation functions for all the hidden units (one layer). Briefly explain what will happen if we initialized all the weights to zero, and all the offset parameters to -1, and ran the stochastic gradient descent method to learn the parameters

(5.6) **(6 points)** We experimented with three different architecture/training combinations:

( ) m hidden units, no regularization

( ) $2m$ hidden units, no regularization

( ) $2m$ hidden units, dropout regularization

Student who carried out the experiments observed that

Model A: training error 0.2, validation error 0.35

Model B: training error 0.25, validation error 0.3

Model C: training error 0.1, validation error 0.4

Please assign the models A,B, and C to their best fitting setups above.