

Q1. Generalization

- (i) Suppose you train a classifier and test it on a held-out validation set. It gets 80% classification accuracy on the training set and 20% classification accuracy on the validation set.

From what problem is your model most likely suffering?

- Underfitting Overfitting

Fill in the bubble next to any measure of the following which could reasonably be expected to improve your classifier's performance on the validation set.

- Add extra features Remove some features

Briefly justify:

- Collect more training data Throw out some training data

Assuming features are outcome counts (k is the Laplace smoothing parameter controlling the number of extra times you "pretend" to have seen an outcome in the training data):

- Increase k Decrease k (assuming $k > 0$ currently)

Assuming your classifier is a Bayes' net:

- Add edges Remove edges

- (ii) Suppose you train a classifier and test it on a held-out validation set. It gets 30% classification accuracy on the training set and 30% classification accuracy on the validation set.

From what problem is your model most likely suffering?

- Underfitting Overfitting

Fill in the bubble next to any measure of the following which could reasonably be expected to improve your classifier's performance on the validation set.

- Add extra features Remove some features

Briefly justify:

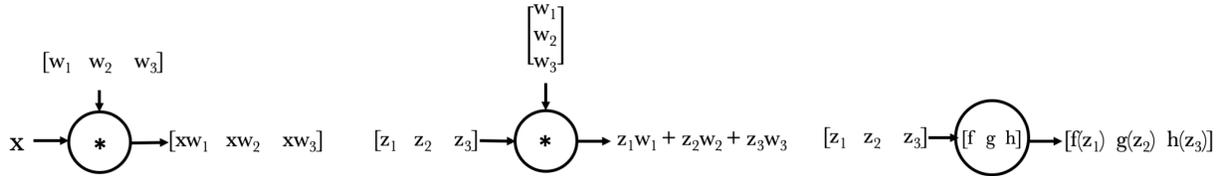
- Collect more training data Throw out some training data

(iii) Your boss provides you with an image dataset in which some of the images contain your company's logo, and others contain competitors' logos. You are tasked to code up a classifier to distinguish your company's logos from competitors' logos. You complete the assignment quickly and even send your boss your code for training the classifier, but your boss is furious. Your boss says that when running your code with images and a random label for each of the images as input, the classifier achieved perfect accuracy on the training set. And this happens for all of the many random labelings that were generated.

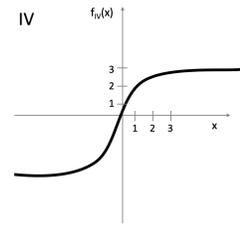
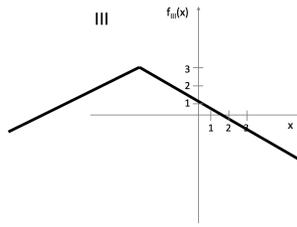
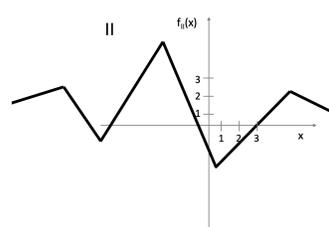
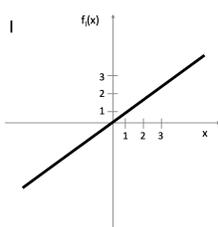
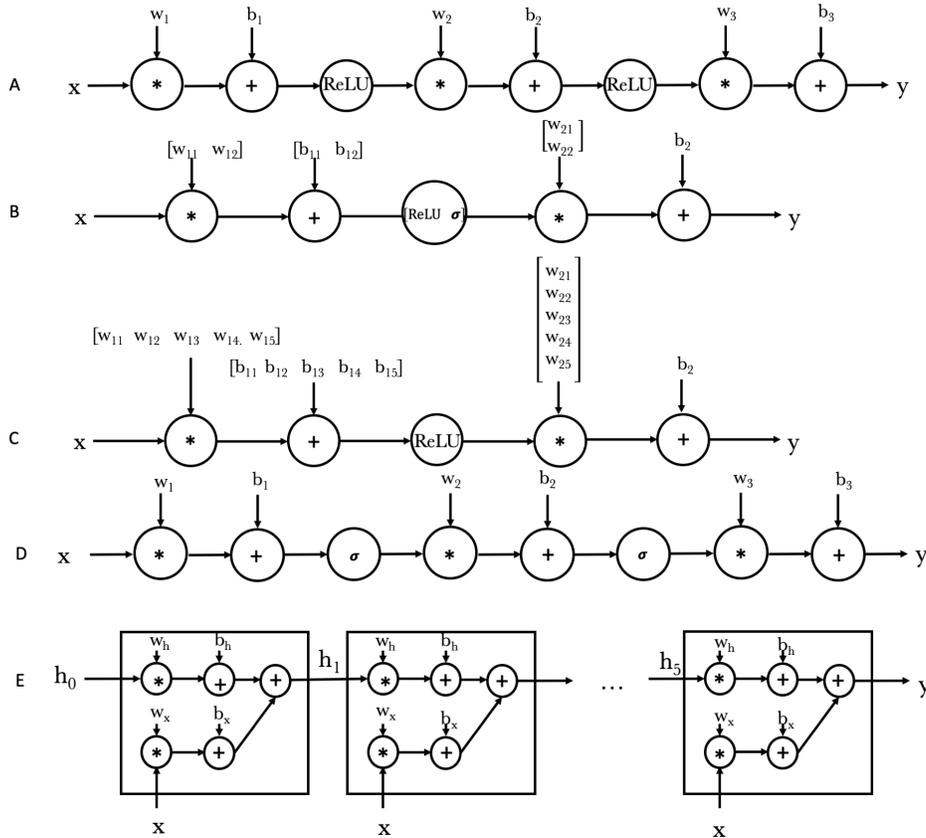
Do you agree that this is a problem? Justify your answer.

Q2. Short Questions

- (a) We are given the following 5 neural networks (NN) architectures. The operation $*$ represents the matrix multiplication operation, $[w_{i1} \dots w_{ik}]$ and $[b_{i1} \dots b_{ik}]$ represents the weights and the biases of the NN, the orientation (vertical and horizontal) its just for consistency in the operations. The term $[\text{ReLU } \sigma]$ in **B** means applying a ReLU activation to the first element of the vector and a sigmoid (σ) activation to the second element. These operations are depicted in the following figures:



Which of the following neural networks can represent each function?



- (i) $f_I(x)$: A B C D E
(ii) $f_{II}(x)$: A B C D E
(iii) $f_{III}(x)$: A B C D E
(iv) $f_{IV}(x)$: A B C D E