

IST 691: Deep Learning in Practice

Homework 01

Instructions: Answer the following questions in no more than one page per question. In addition to the accuracy of your responses, the clarity, coherence, and concision of your writing are critical factors to earning full credit for this assignment. [Cite](#) any source you use outside of lecture notes. This includes the textbook. Reproducing or even paraphrasing an answer from a generative AI tool, such as ChatGPT, is not allowed on this assignment. Submit your responses in single Word or pdf document.

1. In traditional machine learning (non-NN methods), we use feature engineering to model complex relationships between observed variables (features) and a target variable (response). When using deep learning methods, should we design and incorporate feature engineering processes? Explain why or why not.

No, we don't need to incorporate feature engineering processes because deep learning creates rules algorithmically with no human intervention. Although deep learning may not need feature engineering processes, it may need more training data to avoid overfitting or overperforming models.

<https://www.linkedin.com/advice/0/how-can-you-decide-when-use-deep-learning-models#:~:text=One%20of%20the%20main%20advantages,to%20train%20and%20generalize%20well.>

2. Explain in words or equations, why we should introduce nonlinearity in neural networks.

This allows the neural network model to adapt with a variety of data and be able to differentiate between outcomes. There are multiple activation functions in a neural network and depending on the problem, the activation function that fits accordingly is used. These activation functions prevent a network collapse.

<https://www.analyticsvidhya.com/blog/2021/04/activation-functions-and-their-derivatives-a-quick-complete-guide/>

3. You are training a deep learning model to predict sentiment of Twitter posts — the model predicts whether a post is “happy” or “sad”. Your model achieves 0.95 accuracy on the dataset you used to train the model. But when you take new posts from Twitter and use your model to predict the sentiment, the model performs much worse. What might have happened? What should you do to improve your model?

This can be explained by overfitting. This is when a model gives good results for training data but not for new data. This could be because the training data was too small or there's too much irrelevant data, etc. This can be prevented with pruning which identifies important features and eliminates irrelevant ones.

<https://aws.amazon.com/what-is/overfitting/>

IST 691: Deep Learning in Practice

Homework 01

4. The MNIST dataset consists of images of dimension 28×28 pixels with one color channel ($28 \times 28 \times 1$), with each image corresponding to a label between 1 and 10. To build a classifier, we implement a multi-layer perceptron model with 3 hidden layers. The first two hidden layers have 100 perceptrons each, and the third hidden layer has 30 perceptrons. Calculate how many weights will be updated for each iteration of gradient descent. Show your work.

First the # of input pixels must be calculated which is just $28 \times 28 = 784$. The input pixels are connected to each perceptron in the first hidden layer giving us $784 \times 100 = 78,400$ weights. Each perceptron in the first hidden layer is connected to each perceptron in the second hidden layer giving us $100 \times 100 = 10,000$ weights. Finally, perceptrons in the second hidden layer are connected to the perceptrons in the third hidden layer giving us $100 \times 30 = 3,000$ weights. Then finally each perceptron in the third hidden layer is connected to each output layer which gives us $30 \times 10 = 300$ weights. Add them all up to get $78,400 + 10,000 + 3,000 + 300 = 91,700$ weights

5. Answer the following questions based on a close reading of this article and possibly additional research (remember to cite your sources).

<https://www.nytimes.com/2023/06/28/technology/facial-recognition-shoplifters-britain.html>

- a) What is the technology being discussed in this article? How does it relate to deep learning?

Facial recognition is being used in this article. Essentially the deep learning methods learn multiple layers of representations that relate to varied levels of abstraction. This forms a hierarchy of concepts which show strong invariance to the facial pose, lighting and expression changes. With a lot of training data, these facial recognitions models can be very effective.

<https://viso.ai/deep-learning/deep-face-recognition/#:~:text=Deep%20learning%20methods%20use%20a,to%20facilitate%20deep%20face%20recognition.>

- b) What are some ethical concerns discussed in the article? Do you share these concerns? Why or why not?

There are concerns about privacy and poorly performing algorithms. I feel as though these concerns may be warranted as if in the case of catching a criminal. If for some reason an innocent person is falsely matched to the face of someone who committed a crime due to a faulty algorithm, then this is a cause for concern.

- c) Are these concerns addressed by the company deploying this technology? How?

Yes, the company has created policies to prevent misidentification and other errors but mistakes happen. They delete any relevant data to prevent reoccurrence of these errors.

IST 691: Deep Learning in Practice

Homework 01

- d) What is your personal view of the way this technology is being used? Explain your reasoning.

I feel as though the idea has merit in terms of using it to apprehend criminals. If the facial recognition isn't 100% accurate which is probably next to impossible, then it should remain an aid in catching criminals and not the main or only source of conviction against someone. I think it's a helpful tool overall though.