IST 691: Deep Learning in Practice

Homework 01

Ximeng Deng
Sep 24, 2023

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.

[Deep learning] requires very little engineering by hand (LeCun, Bengio, and Hinton 2015). This is because "deep-learning methods are representation-learning methods with multiple levels of representation", and "representation learning is a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification" (LeCun, Bengio, and Hinton 2015). I believe that deep learning is excellent at automatic feature extraction, but there remain instances where manual feature engineering is invaluable.

To begin, a sufficiently deep, well-parameterized neural network would pick up the right features only when "the information is in the data" (StackExchange, 2018). In certain NLP scenarios, semantics are not from data but from prior human knowledge, so features constructed using prior knowledge can be helpful. Moreover, feature engineering can be a critical element of practically deploying deep learning applications at the "edge" where resource constraints might limit the ability to use the entire dataset. An example is streaming sensor data with sampling rates in the hundreds or thousands of Hz (Patrick M.).

Additionally, noisy data requires manual preprocessing, and feature selection can enhance training efficiency. If a "sufficiently deep" network is intractably huge, making model training can be expensive, and we may need to deploy the network in a resource-constrained environment. Feature selection addresses the challenge of overwhelming irrelevant features that may obscure the signal, and it can significantly reduce the number of parameters in the model (StackExchange, 2018).

Furthermore, while deep learning models may be able to automatically generate features, it often compromises interpretability. Many real-world business problems require interpretability, with some even legally binding. Feature engineering allows us to create and select features that are more meaningful and interpretable to humans (Berg and McLendon, 2023).

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

Nonlinearities, such as applying an activation function to each neuron in a neural network, can prevent the network from collapsing. Without nonlinearity, the neurons in a network combine, or collapse, into the equivalent of a single neuron. And one neuron has very little computational power. We use the case and equation from the textbook to see how a network collapses when it doesn't have activation functions (Glassner, 2021)

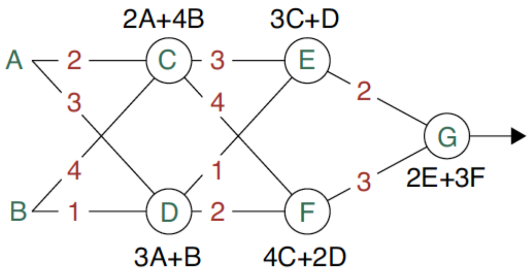


Figure 13-15: Each neuron is labeled with the value of its output.

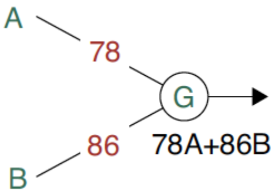


Figure 13-16: This network's output is exactly the same as the output in Figure 13-14.

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?

Our model/network is overfitting. It memorizes the patterns of data in our train set and cannot perform well on the new data that it never seen before. Overfitting is a natural result of training for too long. The problem is that the network learns the training data so well that it becomes tuned to just that data and performs poorly on new data once it's released (Glassner 2021).

Validation and Early Stopping: We use a validation set to estimate a system's generalization error after each epoch during training. Early stopping ends learning when the validation error starts rising. In other words, we can catch when the rules get too specific and stop the learning process at that moment (Glassner 2021).

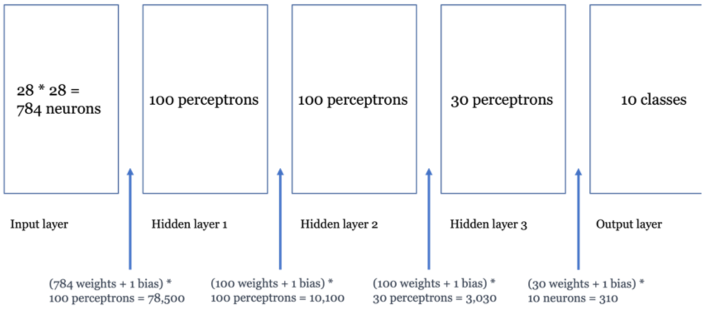
Regularization: Apply regularization techniques such as dropout or batchnorm during model training. Regularization methods delay the onset of overfitting, so we can train longer and continue to push down both training and validation errors.

Dropout delays overfitting by preventing any neurons from overspecializing and dominating. By doing this, the specialized neuron is freed up to perform a more generally useful task, and we've delayed the onset of overfitting. Dropout helps us put off overfitting by spreading around the learning among all the neurons (Glassner 2021).

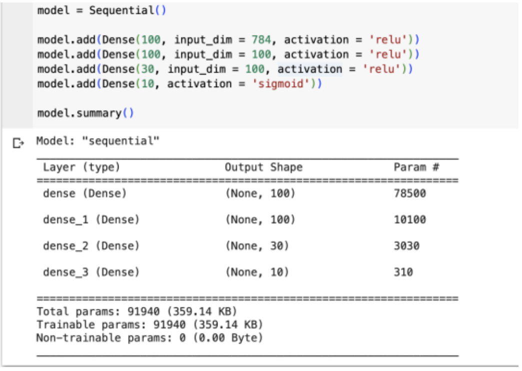
Batchnorm moves the neuron outputs into a small range near 0. We apply batchnorm before the activation function so that the modified values will fall in the region of the activation function where they are affected the most. In this way, we're less prone to seeing any neuron learning one specific detail and producing a huge output that swamps all the other neurons, and thus we are able to delay the onset of overfitting (Glassner 2021).

Collect more recent data and retrain: The training set of posts might be too old. Language, especially on Twitter, changes over time. We can find more data and more recent data to retrain our model, selecting appropriate features or reducing excessive features.

4. The MNIST dataset consists of images of dimension 28x28 pixels with one color channel (28x28x1), 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.


28 * 28 =
784 neurons
100 perceptrons
100 perceptrons
30 perceptrons
10 classes
Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Output layer
(784 weights + 1 bias) * 100 perceptrons = 78,500
(100 weights + 1 bias) * 100 perceptrons = 10,100
(100 weights + 1 bias) * 30 perceptrons = 3,030
(30 weights + 1 bias) * 10 neurons = 310

78,500 + 10,100 + 3,030 + 310 = 91,940



```
model = Sequential()

model.add(Dense(100, input_dim = 784, activation = 'relu'))
model.add(Dense(100, input_dim = 100, activation = 'relu'))
model.add(Dense(30, input_dim = 100, activation = 'relu'))
model.add(Dense(10, activation = 'sigmoid'))

model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 100)               78500

 dense_1 (Dense)             (None, 100)               10100

 dense_2 (Dense)             (None, 30)                3030

 dense_3 (Dense)             (None, 10)                310

=================================================================
Total params: 91940 (359.14 KB)
Trainable params: 91940 (359.14 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

91,940 weights (including bias) will be updated for each iteration of gradient descent.

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?

The technology being discussed in this article is Facial Recognition Technology. Retail store security officers pulled up security camera footage, saved the photo of the thief’s face, and upload it to a facial recognition program, Facewatch, which is associated with other stores’ surveillance video systems. Facewatch is trained on millions of pictures and videos, the system reads the biometric information of a face as the person walks into a shop and check it against a database of flagged people. Once this criminal’s face reappears in another store’s surveillance video and is recognized by this Facewatch program, it will generate a Match notification to store security officer.

Facial recognition technology often relies on deep learning algorithms, particularly Convolutional Neural Networks (CNNs), to perform accurate and efficient identification or verification of faces in images or video streams. These networks learn to extract hierarchical features, starting from low-level features like edges and corners, and gradually building up to high-level features like facial contours and expressions (RecFaces).

The algorithms perform three main tasks: detect faces in an image, video, or real-time stream; calculate a mathematical model of a face; compare models to training sets or databases to identify or verify a person (RecFaces).

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

Concerns:

- “Individuals have little way of knowing they are on the watchlist or how to appeal.”
- “Privacy concerns and poorly performing algorithms in the past.” (Fraser Sampson, Britain’s biometrics and surveillance camera commissioner, who advises the government on policy)
- “The technology did not leave much room for human discretion.” (The lady who may have walked out after not realizing that her debit card payment failed to go through at a self-checkout kiosk)
- “Its deployment to prevent petty crime might be illegal under British privacy law, which requires that biometric technologies have a ‘substantial public interest.’” (Civil liberties groups)
- “Example of misidentification. … a man was confused for his identical twin, who had shoplifted.” (Liam Ardern)

Yes, I totally agree that these concerns and issues surrounding facial recognition technology need urgent resolution. It’s unacceptable for individuals who are falsely accused of criminal behavior or misidentified as someone else to have no means to appeal.

Should this technology, still in need of refinement, be deployed, it’s crucial for the facial recognition companies (e.g., Facewatch) to bear legal and financial responsibility for instances of misrecognition. More importantly, businesses that employ this technology should undergo training - potentially provided by Facewatch - to allow for a measure of human discretion and customer negotiation in cases of false identification.

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

Yes, Facewatch is making efforts to address some of these issues.

(1) First off, photos of individuals labeled as shoplifters and problematic customers remain in the database for a year before being deleted, which is a reasonable approach. (2) Secondly, Facewatch's system currently employs a "super recognizer" for human-assisted double verification, effectively reducing the number of AI-based errors. (3) Thirdly, Mr. Gordon apologized for the Bristol incident, stating, "If this occurs, we acknowledge our mistake, apologize, delete any relevant data to prevent reoccurrence, and offer proportionate compensation." (4) Additionally, Stephen Bonner mentioned that Facewatch is changing its policies to include more signage in stores. I agree with this approach, as it parallels the common practice of posting "This Property is Protected by Electronic Surveillance" signs. (5) Lastly, Facewatch will only share information about serious and violent offenders among stores and will only send out alerts about repeat offenders. This means that people will not be added to the watchlist after a single minor offense, as happened to the woman in Bristol. "This reduces the amount of personal data held, lowers the chances of individuals being unfairly added to such lists, and increases accuracy," Mr. Bonner added.

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

I think facial recognition technology has its place in retail stores. Its primary purpose is to protect stores' profits and prevent shoplifting. We shouldn't encourage bad behavior, and if shoplifters aren't held accountable, the costs due to the theft that happens every week pile up, which could lead to potential layoffs, or higher prices for customers. When we talk about human rights, it isn't just privacy we need to consider. Technology can also play a role in safeguarding other aspects of our rights.

It is true that new technology can have teething problems. But avoiding it because of potential privacy issues isn't the solution. Only when we deploy the technology can we identify and resolve these issues as they arise, refining the technology as we go. Otherwise, it's like taking a step back in time.

And if we question the ethics of facial recognition, shouldn't we also discuss surveillance cameras that have been around for years? Where surveillance cameras are legal, having surveillance footage without optimizing its utility seems wasteful. Facial recognition can enhance the effectiveness of these existing systems.

Another point to consider is the trust, or lack thereof, in the police. If the police can't or won't chase down every petty crime, then using technology to deter and detect crimes becomes even more critical. It becomes a way for stores to proactively handle security, especially if they can't rely solely on law enforcement.

Citation:

LeCun, Y., Bengio, Y. & Hinton, G. "Deep learning." *Nature* **521**, 436–444 (2015). https://doi.org/10.1038/nature14539.

"Why do neural networks need feature selection / engineering?" *StackExchange*, Modified May 31, 2018, https://stats.stackexchange.com/questions/349155/why-do-neural-networks-need-feature-selection-engineering. Accessed 24 Sep 2023.

M., Patrick. "How do you use data augmentation and feature engineering to enhance your deep learning model?" *LinkedIn*, https://www.linkedin.com/advice/3/how-do-you-use-data-augmentation-feature-engineering#:~:text=Feature%20engineering%20can%20help%20you,domain%20knowledge%20and%20prior%20information. Accessed 24 Sep 2023.

Berg, Leah, and Ray McLendon. "Is Feature Engineering Dead?" *Medium*, Towards Data Science, 10 Jan. 2023, https://towardsdatascience.com/is-feature-engineering-dead-203e6a9e5751. Accessed 12 Sep. 2023.

Glassner, Andrew. "Deep Learning – A Visual Approach." No Starch Press, 2021, pp.188, 197, 203, 329-330, 422-423.

"Understanding Facial Recognition Algorithms." *RECFACES*, https://recfaces.com/articles/facial-recognition-algorithms. Accessed 16 Sep. 2023.