CS 6762
Fall 2023
SP, ML, FC Homework 3

Answer all questions. Work alone. Total is 100 points. Type or write neatly.

1. Describe entropy, information gain and the relationship between the two. (5 points)

Entropy (H(X)) is the measure of uncertainty of disorder in a random variable or set of items. Entropy is highest when the probability of all outcomes are equal and lowest when some are more probable than others.

Information Gain (IG) is a measure used in decision trees to quantify how much "information" or entropy is lost when conditioning on another target variable. IG = H(Y) – H(Y|X). This essentially provides you with the information gained from Y given we already knew X.

2. Assume an urn has 3 red balls, 2 green balls, and 1 white ball. Compute the entropy. (5 points)

$$P_{red} = \frac{3}{6}, P_{green} = \frac{2}{6}, P_{white} = \frac{1}{6}$$
$$H(X) = -\sum p(x_i) \log_2 p(x_i)$$
$$H(X) = -(0.5 \log_2 0.5 + 0.333 \log_2 0.333 + 0.167 \log_2 0.167)$$
$$H(X) = -(-0.5 + -0.528 + -0.432)$$
$$H(X) = 1.46$$

3. What is a random forest classifier and state its advantages over a decision tree. (5 points)

A random forest classifier is an ensemble learning technique that combines the output of multiple decision trees to generate a prediction. This is the mode of the classes in a classification context and the average of the values in a regression context. The trees in the forest are powered by a concept known as bagging (bootstrap aggregating) where random subsets of both the rows and columns of the original dataset are sampled with replacement to produce varied training sets for each tree.

The advantages over a single decision tree are numerous including:

1. Reduced Chance for Overfitting
    a. Predictions from a wide variety of trees being averaged together heavily reduces variance and as a result overfitting
2. Higher Accuracy
    a. The model benefits from the ensemble effect where multiple weak learners combine form a stronger overall model.
3. Feature Importance Score Calculation

    a. A novel evaluation metric for individual features, the feature importance, can be calculated for a given random forest to see how impactful each feature is.
4. Robust Against Noise
    a. Each tree being trained on a unique subset of the rows and columns of the data allows robustness against noise and data outliers.

4. What is a sigmoid function? A tanh function? A softmax? (5 points)

A sigmoid function squashes an input value into a 0-1 range. It is typically deployed in a binary classification context in the output layer.

A tanh function squashes an input value into a -1 – 1 range. It is typically used in neural networks within hidden layers where you want to allow both positive and negative values.

A softmax function squashes a vector of values to sum to 1 and is typically used as the output layer of a neural network that performs multi-class classification.

All of these are examples of activation functions that introduce non-linearity into traditionally linear machine learning models. This allows those models to learn more complex relationships in the input data.

5. Describe in 1 or 2 paragraphs how to train a NN (look this up on the Internet and/or watch some YouTube videos). (10 points)

Training a neural network (NN) involves the process of adjusting the model's weights and biases to minimize the error between its predictions and the actual target values. This is typically done through an iterative process using a method called backpropagation combined with an optimization algorithm like stochastic gradient descent (SGD) or one of its variants.

This often takes an arbitrary number of iterations and epochs (full passes over the entire data set) until an evaluation metric reaches a value that is satisfactory for the operator training the model. Once this has been reached the model is ready for prediction on new/unseen data.

6. Briefly describe 2 applications from any of the smart health, smart cities, or autonomous systems application areas where you can/should use a DNN. (10 points)

Smart Health -> Medical Image Analysis: DNN’s can often be used in medical image analysis to identify tumors, fractures, or any other abnormalities that may be difficult to discern. Convolutional Neural Networks are a typical DNN applications for medical images like X-Rays and CT-Scans. These models can sometimes exceed human performance motivating their use.

Smart Cities -> Traffic Prediction: DNN’s can be used in smart cities to predict and optimize traffic light patterns so as to reduce congestion in city streets. These models are trained on past historical data and predicted on fresh traffic data to model and predict traffic in cities.

7. CNNs were originally developed for use with images. Explain how to use CNNs for 2 problems that are NOT related to images. For this, explain the inputs and eventual outputs of the CNN. (20 points)

Text Classification:

- Input: Text matrices where each column is a feature of the word embeddings and each row is an individual textual "token" usually representing a single word. Produces a 2D matrix that a CNN's kernel can convolve over.
- CNN Layers: CNN layers slide filters over the input sequence and likely capture local patterns and word sequences that are predictive towards the final output like "not good" or "very happy".
- Output: Binary or Multi-Class Classification of input sequence as positive/negative or an arbitrary number of possible classes.

Time Series Data – Stock Price Prediction:

- Input: Stock data where each row represents a time step and each column is a piece of data at that time step. Produces a 2D matrix that a CNN's kernel can convolve over.
- CNN Layers: CNN layers will capture local temporal patterns like upward or downward trends of the stock price.
- Output: Binary classification of the stock going up/down or a literal number representing the next predicted price of the stock.

8. Describe the main gates in an LSTM cell, including their purpose, their inputs and outputs, and what is inside the gate. (10 points)

1. Forget Gate: The forget gate controls how much of the previous cell will be discarded (forgotten) and is controlled by a weight matrix and a bias term. It essentially allows the LSTM to forget irrelevant parts of past information.
2. Input Gate: This gate controls how much of the new information presented to this LSTM node will actually be added to the nodes cell state. This contains a sigmoid activated input filter as well as a tanh activated candidate value matrix that are multiplied together and learned independently allowing the model to selectively incorporate new information.
3. Output Gate: This determines how much of the current cell state should be exposed as output of this cell for the next cell. It is also controlled by a sigmoid activated output matrix that determines how much of the information to reveal which is multiplied by a tanh activated cell state and represents that cells hidden state.

9. What are the key issues in developing DNNs for execution on small devices? (5 points)

The main issue is typically the size of these models combined with the small size and relative computational power of these devices. Low computational power as well as low ram and storage space create for a difficult model hosting scenario. These problems can be alleviated using techniques like model pruning but there isn't necessarily a solution, typically a specific model is chosen that fits more readily on smaller hardware.

~~10. In the SparseSep system we studied, what are the inputs and outputs to its Layer Compression Compiler? (5 points)~~

11. Describe how ML can be used in autonomous systems and what are its challenges? (10 points)

Machine learning (ML) plays a crucial role in autonomous systems, helping them with tasks like recognizing objects, making decisions, and navigating through different environments. ML processes sensor data using techniques like computer vision and reinforcement learning to help the system understand its surroundings. It's also used for mapping, localization, and predicting when the system might need maintenance. However, there are some big challenges like making sure the system works in real-time, stays safe, and can handle different kinds of environments. ML also needs a lot of data and making sure its decisions are understandable and ethical is tough. Solving these issues is important for making ML work well in autonomous systems.

12. Describe how ML can used in smart health CPS systems and what are its challenges? (10 points)

Machine learning is useful in smart health CPS systems because it helps analyze patient data, track vital signs, and even assist with diagnosing diseases. For example, it can take data from wearable devices to spot health problems early or predict what might happen to a patient based on past data. It also helps with personalized treatments by analyzing a patient's medical history. But, there are some challenges like making sure patient data stays private, dealing with the huge amount of data, and making sure the predictions are reliable. Also, it's hard to understand how ML makes decisions with regard to explainability, which can make doctors hesitant to trust it.

Note: Questions 11 and 12 are not the same as question 7 which is more related to the mechanics of the CNN, although there may be some overlap.