ML ASSIGNMENT 1: Classification

Hitesh Sharma, Aditya Vyas, Atharva Pawar

BERT Models for Sentiment Analysis and Emotion Recognition

This presentation explores the use of BERT models for sentiment analysis and emotion recognition of Twitter data. It discusses the advantages of social media data, learning-based approaches, and the challenges of deep learning models.

Domain : Classification
Link to Paper : https://ceur-ws.org/Vol-2841/DARLI-AP_17.pdf

Link to implementation :
https://colab.research.google.com/drive/14aLspmK2MbTUmECW-GvT5M4kQBaAvbKn?usp=sharing

Introduction

1. The paper investigates the use of BERT models for sentiment analysis and emotion recognition of Twitter data.
2. Social media data is considered authentic and provides valuable information about user opinions on various topics and products.
3. Learning-based approaches have shown better performance in sentiment analysis and emotion recognition than lexicon-based approaches.
4. Deep learning models require large corpora of labeled data, which can be scarce and expensive to build.

Pre-trained BERT Models

1. Pre-trained models that only need a fine-tuning phase with a smaller dataset have been widely used.
2. Many neural networks composed of a task-agnostic pre-trained word embedding layer and a task-specific neural architecture have been proposed.
3. Recent architectures based on Transformer have shown further room for improvement.
4. The authors define two separate classifiers for sentiment analysis and emotion recognition tasks.
5. The models achieve an accuracy of 0.92 and 0.90 on sentiment analysis and emotion recognition, respectively.
6. The authors use a dataset of 1.6 million tweets for training and testing their models.

Experimental Results

1. The authors use a pre-trained BERT model and fine-tune it on their dataset.
2. The authors use a combination of attention mechanisms and pooling strategies to extract features from the BERT model.
3. The authors use a softmax layer for classification.
4. The authors compare their models to several baselines, including SVM, Naive Bayes, and LSTM models.

Comparison and Conclusions

1. The authors find that their models outperform the baselines in terms of accuracy and F1 score.
2. The authors also perform an ablation study to analyze the contribution of different components to the overall performance of the models.
3. The authors find that the attention mechanism and pooling strategy are important for achieving high accuracy.
4. The authors also analyze the performance of their models on different subsets of the dataset, such as tweets with hashtags or mentions.
5. The authors find that their models perform well on all subsets of the dataset.
6. The authors conclude that their BERT-based models are effective for sentiment analysis and emotion recognition of Twitter data and can be used in real-world scenarios such as marketing or political campaigns. They also note that future research can explore the use of BERT models for other social media platforms and languages.

Results

https://github.com/capstone-project-SECURIX/ml-projects/tree/main/Ass1-Classification%20Algorithm

Positive Test

```
[ ]  # Example usage:
     test_sentence = "I really enjoyed that movie" # Positive
     # test_sentence = "It was a bad movie" # Negative
     result = predict_sentiment(test_sentence)
     print(f"Sentence: {test_sentence}")
     print(f"Sentiment: {result}")
```

```
Sentence: I really enjoyed that movie
Sentiment: Positive
```

Negative Test

```
[ ]  # Example usage:
     # test_sentence = "I really enjoyed that movie" # Positive
     test_sentence = "It was a bad movie" # Negative
     result = predict_sentiment(test_sentence)
     print(f"Sentence: {test_sentence}")
     print(f"Sentiment: {result}")
```

```
Sentence: It was a bad movie
Sentiment: Negative
```

Housing Price Prediction Based on Multiple Linear Regression

Housing Price Prediction Based on Multiple Linear Regression is a statistical technique used to predict the price of a house based on various factors. The dependent variable in this case is the house price, and the independent or explanatory variables could include factors like the size of the house, the location, the age of the house, and others. The model estimates the relationship between these independent variables and the dependent variable, providing insights into how different factors affect the housing price. This technique is particularly useful in real estate where understanding the factors influencing house prices can aid in making informed decisions about property investments.

Domain : Classification
Link to Paper : https://www.hindawi.com/journals/sp/2021/7678931/

Link to implementation :
https://colab.research.google.com/drive/1vu9LeRKrns_Nbz7UQagrqHQ15pmEQmLL?usp=sharing

Introduction

- Housing Price Prediction Based on Multiple Linear Regression is a statistical method that uses multiple independent variables to predict the dependent variable, which in this case is the price of a house.
- The model is built by estimating the relationship between these independent variables (like the size of the house, location, age of the house, etc.) and the dependent variable (the house price).
- This technique is particularly useful in real estate for understanding the factors influencing house prices and making informed decisions about property investments

Housing Price Prediction

This presentation explores the complex task of housing price prediction using statistical techniques.

Multiple Linear Regression Model

Model trained and tested using Boston dataset

Parameters optimized with gradient descent optimizer

Comparison of prediction results with real values

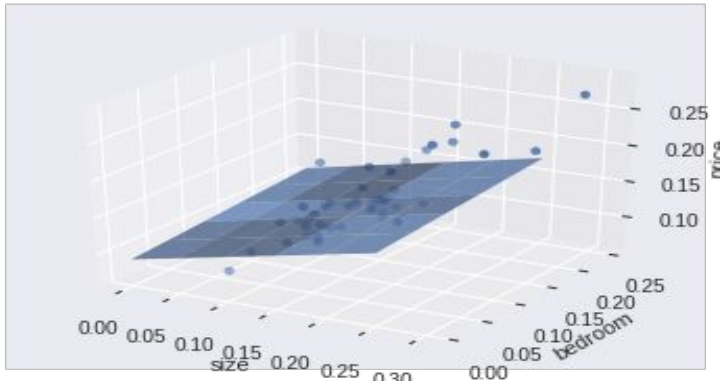
Comparison and Conclusions

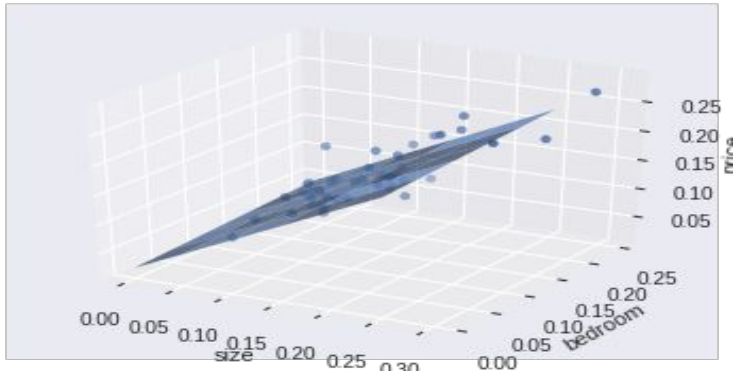
- Comparison: Multiple Linear Regression provides a more accurate prediction of housing prices compared to simple linear regression. This is because it considers multiple independent variables, allowing for a more nuanced understanding of the factors influencing housing prices.

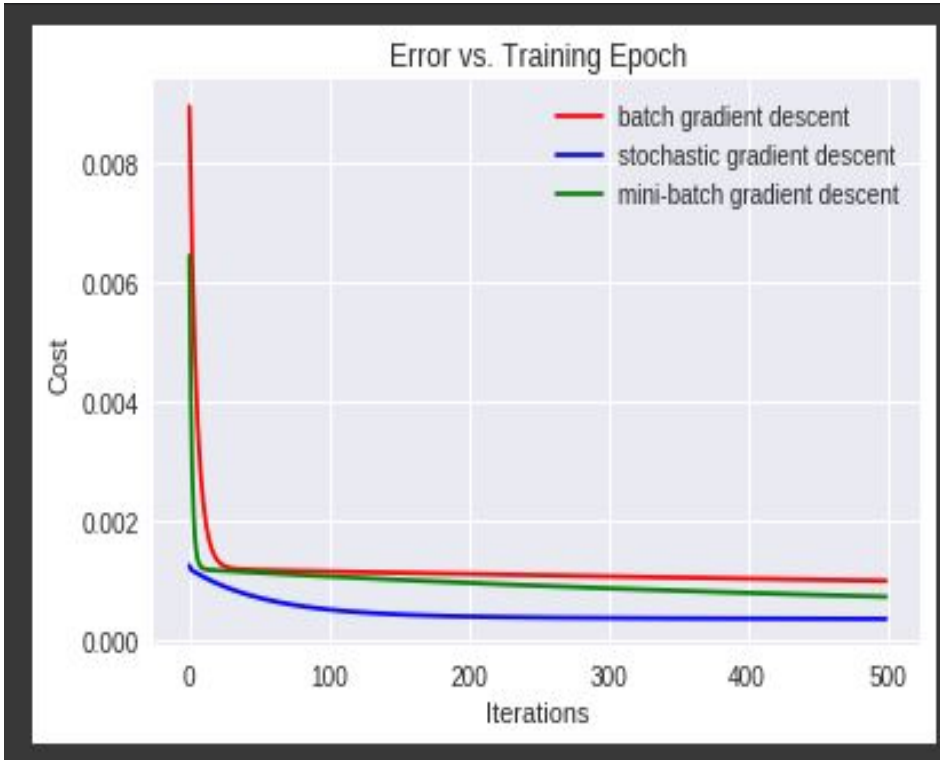
- Conclusion: The model's results change drastically when introducing new variables, indicating that a more robust model is required. The multiple linear regression model is better than the simple linear model as it is statistically significant at the 5% threshold.

  The modeling process is iterative, involving testing, analyzing, failing, and testing some more. This approach is crucial in understanding housing prices, which is a complex problem with many predictor variables.

Results


0.25
0.20
0.15
0.10
0.00 0.05 0.10 0.15 0.20 0.25 0.30
size
0.00 0.05 0.10 0.15 0.20 0.25
bedroom


0.25
0.20
0.15
0.10
0.05
0.00 0.05 0.10 0.15 0.20 0.25 0.30
size
0.00 0.05 0.10 0.15 0.20 0.25
bedroom


Error vs. Training Epoch
batch gradient descent
stochastic gradient descent
mini-batch gradient descent
Cost
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
Iterations

Thank you for your time and attention 🙂