Security and Privacy of Machine Learning-Spring 2020 Research Project Proposal
Adversarial Examples on Tabular Data

Zuhal Altundal
Syracuse University
College of Engineering &
Computer Science
zaltunda@syr.edu

ABSTRACT

In recent years, Artificial Intelligence technology be- came most popular used technology and is developing very fast. And AI technology comes with its own security problems. One of these security problems method is adversarial examples, that usually work based on the design of attacks and defenses with an active focus on the image domain. Furthermore, many machines learning classifiers in the financial industry like as to predict credit rate of the person. For example, a customer wants to apply a loan and banker is using the machine learning to make a decision regarding the customer provided information.

KEYWORDS

Adversarial Examples, Security Machine Learning, Tabular Data Gradient Boost Classifier

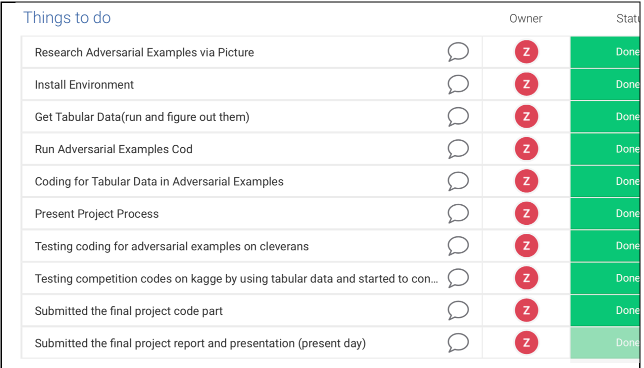
Introduction

In this project paper, we propose to do adversarial at- tack on the tabular data that is used for making a decision whether the loan is accepted for customer who has already applied or not. The adversarial examples for image data satisfy pixels that are defined as integers between 0 and 255.

These properties need to get attack for tabular data. We need to fit our imperceptibility data (fake data) into tabular data. For example, phone bill information might be forced into a positive number or removed depending on the problem that cause of rejecting the applied loan. Images and tabular data have not the same perceptibly concepts. We cannot use same formalization that we use adversarial attack on the images, on the tabular data. We need to create new methods to make adversarial attack on tabular data. For adversarial attack on tabular data, we provide in this project to manipulate the data information that is credit rate information to change credit score of people who has not enough score for taking loan.

Software Requirements

- Python
- PyTorch is an open-source machine learning library for Python, based on Torch, used for applications such as natural language processing



Things to do	Owner	Stat
Research Adversarial Examples via Picture	Z	Done
Install Environment	Z	Done
Get Tabular Data(run and figure out them)	Z	Done
Run Adversarial Examples Cod	Z	Done
Coding for Tabular Data in Adversarial Examples	Z	Done
Present Project Process	Z	Done
Testing coding for adversarial examples on cleverans	Z	Done
Testing competition codes on kagge by using tabular data and started to con...	Z	Done
Submitted the final project code part	Z	Done
Submitted the final project report and presentation (present day)	Z	Done

Figure-1 Things to do through project

Data Set

- Credit Kaggle Competition

Data Descriptions

Variable Name	Description	Type
SeriousDlqin2yrs	**Person experienced 90 days past due delinquency or worse**	**Y/N**
RevolvingUtilizationOfUnsecuredLines	Total balance on credit cards and personal lines of credit except real estate and no installment debt like	percentage

	car loans divided by the sum of credit limits	
age	Age of borrower in years	integer
NumberOfTime30-59DaysPastDueNotWorse	Number of times borrower has been 30-59 days past due but no worse in the last 2 years.	integer
DebtRatio	Monthly debt payments, alimony,living costs divided by monthy gross income	percentage
MonthlyIncome	Monthly income	real
NumberOfOpenCreditLinesAndLoans	Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)	integer
NumberOfTimes90DaysLate	Number of times borrower has been 90 days or more past due.	integer
NumberRealEstateLoansOrLines	Number of mortgage and real estate loans including home equity lines of credit	integer
NumberOfTime60-89DaysPastDueNotWorse	Number of times borrower has been 60-89 days past due but no worse in the last 2 years.	integer
NumberOfDependents	Number of dependents in family excluding themselves (spouse, children etc.)	integer

After we finished research about adversarial examples, we started to install requirement framework, and library for running a source code of adversarial example via figures. What are the requirement for it;

• Examine structure of Cleverhans that is a Python library to benchmark machine learning systems' vulnerability to adversarial examples

• Install Python 3.5

• Install TensorFlow 1.12-require to install virtual environment in Ubuntu

• OS will be Ubuntu 14.04.5 LTS (Trusty Tahr)

Then we started to set up Cleverhans by installing above requirements. But when we installed the virtual environment we got some error that we work on it to fix it. If we would install successfully the requirements that require in work environment for our project, we will start to run source code in Github, we plan to finish figure out the source code in 15th March as we mentioned the due date in timeline sheet in Figure1.

We are going to use as dataset" give me some credit" that downloaded already from website https://www.kaggle.com. These dataset (information in Figure2) will be our tabular data that we will use in our project. Dataset configuration task's due date is in 18th March as we mentioned in Figure1. If we would finish our schedule for our project until 18th March successfully, we are going to start coding in python to finish last task in our project that its due date is in 30th March.

Imperceptible Attacks on Tabular Data

On this paper, we need to examine adversarial examples code how is working. We used the cleverans example that we include link on the reference list. After we worked on the adversarial example codes, we can say that adversarial examples are using usually based on the images data. But the main idea is feeding the data by using fake structures. When we use this idea on tabular data like as credit card score data, we need to know there are some different perspective for tabular data from images. First of all, tabular data features are not interchangeable like as pixels, and the tabular data are less readable and expert knowledge is required instead of the images. Image classes can be appeared by anyone.

That is why, attacker should avoid to modificative the subset of features on tabular data, and so we tried to train tabular data by using argmin that arguments of the minimum of any features on dataset. This argmin function can develop for any specific results. We know that credit score prediction mechanisms are usually running by using median functions. We just found an argmin results inside of the median calculation and change the data with our argmin data for training the tabular data. This actually change some minimum structures of results.

When we used it we just added this argmin part inside of the competition code on Kaggle.

```
def rmDataAndPutArgmin(train_data, first = 
35, second = 40):
    New = []
    argminx = train_data.argmin()
    for val in train_data:
        if ((val==first)| (val == second)):
            New.append(argminx)
        else:
            New.append(val)

    return New
```

```
//Remove any specific data and put
argmin of it. The condition can
develop whatever we want to get or
delete or get result of the function
```

we used gradient boost classification algorithm for calculation of prediction and represent the probability that a random positive side is positioned to the right of a random negative side

Conclusion

In this project, we tried to study adversarial examples on tabular data. after this project based on our knowledge, images and tabular data don’t share same perceptibility concepts. To generate similar of adversarial examples we just tried to change minimize of any specific data before training and used gradient boos classifier show us the AUC after we put our data. On our opinion, this type of attack may more challenge than attack of adversarial examples on images data because it depends on the expert eye on data.

REFERENCES

- Vincent Ballet, Xavier Renard, Jonathan Aigrain, Thibault Laugel, Pascal Frossard, Marcin Detyniecki https://arxiv.org/pdf/1911.03274.pdf
- freeCodeCamp.org
- https://github.com/max-fitzpatrick/Credit-scoring-model/blob/master/CREDIT_SCORING_NOTEBOOK.ipynb
- https://www.kaggle.com/simonpfish/comp-stats-group-data-project-final
- https://www.kaggle.com/brycecf/give-me-some-credit-dataset
- https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab
- http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf
- https://arxiv.org/pdf/1412.6572.pdf
- https://github.com/tensorflow/cleverhans/tree/master/cleverhanstutorials
- https://github.com/gabmars/Give-me-some-credit-Kaggle-/blob/master/GiveMeSomeCredit.py