**Project Report - CS539 Machine Learning - Spring 2017**

**Prof. Carolina Ruiz, TA Ahmedul Kabir**

**Students: \_\_\_\_Fangling Zhang\_\_\_\_\_\_ and \_\_\_Fangyu Lin\_\_\_\_\_\_\_\_\_\_**

**I. General Project Description (at most 2 pages)**

1. Selected Machine Learning Method to Focus on:

**Artificial Neural Networks and Deep Learning**

1. Selected Advanced Topic Related to your Focus Machine Learning Method:
   1. Topic name: **Image Classification with Deep Convolutional Neural Networks and Pre-trained Networks.**
   2. How is it related to your focus machine learning method?

**To get high accuracy of image Classification, we need to construct many layers of neural networks. Getting the trained models usually takes us much time. Pre-trained Deep Networks provide models already trained by others, which can be used to obtain features and test our image data much faster. In our case, we use pre-trained AlexNet and cifar10Net to classify our images**

* 1. Scientific sources or textbooks you used to investigate this topic:

**(1) Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks.**

**(2)** [**Jason Brownlee**](http://machinelearningmastery.com/author/jasonb/)**. How To Improve Deep Learning Performance.**

* 1. Summary of what you learned about this topic:

1. **How to construct a deep convolutional neural networks to classify images and how to choose parameters.**
2. **What’s AlexNet and how to use it for fast image recognition.**
3. **Why AlexNet can provide high accuracy of image detection.**
4. **AGG and other pre-trained networks.**
5. **Use DCNN and pretrained networks to classify 2 classes and multiclass images.**
6. Dataset
   1. Name: [**CIFAR-10 datase**](https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz)**t**
   2. Where found (location):

[**https://www.cs.toronto.edu/~kriz/cifar.html**](https://www.cs.toronto.edu/~kriz/cifar.html)

* 1. Dataset Description: [Including number of data instances, number of attributes, target attribute, percentage of missing values, and any other salient characteristics of the dataset]

1. **60000 32x32 images data in 10 classes. In our project, we choose bird images (6000) and cat images (6000) as binary classification. Also, we will use the 10 classes original data as multiple classification.**
2. **3072 Attributes. The first 1024 (=32\*32) entries contain the red channel values, the next 1024 the green, and the final 1024 the blue.**
3. **Target attribute: Binary classification: ‘bird’, ‘cat’. Multiple classification: 'airplane','automobile','bird','cat','deer','dog','frog','horse','ship','truck'**.
4. Initial data preprocessing, if any: **We select bird and cat images from original images which mix 10 classes of images together. Reshape the images data from 2-Dimensions to 4-D so that it can be used by CNN layer.**
5. Purpose and goals of your experimentation. What do you want to learn from this dataset using your selected machine learning method? Why is this method particularly appropriate for this purpose?

**Goal: Develop the accuracy and speed of images classification. We want to classify images correctly and fastly. it is one of the most efficient methods for extracting critical features for non-trivial tasks. Images have many features so we need to impose diminishing small features to recognize images correctly and quickly.**

1. Experimental protocol. Describe here details of experimentation that you plan to perform:
   1. k-fold Cross-validation. Yes? No? (if so, why not?) k=?

**Yes. K=3.**

* 1. Experimentation with and without pre-processing. What pre-processing techniques will you use?

**Yes, We will try whitening transformation.**

* 1. Experiments with and without dimensionality reduction. What dimensionality reduction techniques will you use?

**Yes, use pre-trained networks (such as AlexNet) to select features.**

* 1. Experiments to construct models with your selected machine learning method. What method parameters do you plan to vary? Will you use any post-processing?

**Method parameters to vary: if use bias, local normalization and dropout layer or not; initial learning rate, initial weight and number of epochs.**

**We will compare the accuracy and time cost of these experiments with tables and figures.**

* 1. Experimentation with the advanced topic you selected related to your machine learning focus.

**Method parameters to vary: feature selecting layer, Mini-Batch size .**

* 1. What other machine learning methods studied in class (and listed in the project webpage) will you use to compare against your focus machine learning method? Include as many as possible.

**Other machine learning methods: Decision tree, Naive Bayes, Logistic classification, SVM.**

* 1. What meta-learning techniques will you use to combine multiple methods / models.

**We will use boosting, bagging and stacking in optimization process of neural networks.**

* 1. What quantitative and qualitative approaches will you use to compare results of the different experiments above?

**Test accuracy of classification.**

**II. Experiments ran on the dataset with your focus machine learning method and results (at most 2 pages)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **model name** | **Brief description of the experiment** | **data pre-processing** | **Time take(s)** | **Test Accuracy** |
| model0 | basic model with all default options in CNN function |  | 2139 | 0.725 |
| model1 | add initial bias to lay 2 to basic model |  | 2070 | 0.750 |
| model2 | add local normalization layers to model 1 |  | 2213 | 0.475 |
| model3 | add dropout layers to model 1 |  | 2367 | 0.740 |
|  | use whitening transformation firstly, then use model 3 | whiten | 2113 | 0.730 |
| the following models are based on model 3, just change initial weight coefficient(w0), initial learning rate(LR) and max epochs(epoch) | | | | |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

[This section should contain: 1. a description of each experiment you ran on the dataset in terms of data pre-processing performed, time taken to run the experiment, values of evaluation metrics you used to analyze the results of the experiment (e.g., time taken to run the experiment, accuracy, ROC area, precision, recall, …), qualitative evaluation of the models you obtained (patters described by the model). 2. your experimentation with the advanced topic related to your focus machine learning method that you investigated for this project.]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **classes number** | **Brief description of the experiment** | **data pre-processing** | **Time take(s)** | **Test Accuracy** |
| 10 | Use pre-trained cifar10Net to classify 10 classes image | feature selection before train and test | 200.0670 | 0.7456 |
| 10 | Use pre-trained AlexNet to classify 10 classes image (learner:linear, feature selecting layer(l):’fc8’, Mini-Batch size(m):32 | feature selection before train and test | 654.9856 | 0.5686 |
| 2 | basic model: Use pre-trained AlexNet to classify 10 classes image (parameter is the same as above model) | feature selection before train and test | 131.1835 | 0.8143 |
| the following models are based on above basic model, just change feature selecting layer(l) and Mini-Batch size(m) | | | | |
|  | l m |  |  |  |
| 2 | |  |  | | --- | --- | | 'conv2' | 8 | | 'conv2' | 16 | | 'conv2' | 32 | | 'conv3' | 8 | | 'conv3' | 16 | | 'conv3' | 32 | | 'conv4' | 8 | | 'conv4' | 16 | | 'conv4' | 32 | | 'conv5' | 8 | | 'conv5' | 16 | | 'conv5' | 32 | | 'fc6' | 8 | | 'fc6' | 16 | | 'fc6' | 32 | | 'fc7' | 8 | | 'fc7' | 16 | | 'fc7' | 32 | | 'fc8' | 8 | | 'fc8' | 16 | | 'fc8' | 32 | | feature selection | 68.4524 | 0.7857 |
| 2 | feature selection | 68.2882 | 0.7571 |
| 2 | feature selection | 68.6478 | 0.6714 |
| 2 | feature selection | 98.2388 | 0.8429 |
| 2 | feature selection | 99.399 | 0.8143 |
| 2 | feature selection | 99.9337 | 0.8143 |
| 2 | feature selection | 116.0749 | 0.7857 |
| 2 | feature selection | 117.7081 | 0.7571 |
| 2 | feature selection | 118.2756 | 0.7857 |
| 2 | feature selection | 128.4025 | 0.6857 |
| 2 | feature selection | 129.8391 | 0.6857 |
| 2 | feature selection | 130.4574 | 0.7286 |
| 2 | feature selection | 130.0392 | 0.8 |
| 2 | feature selection | 130.5892 | 0.8143 |
| 2 | feature selection | 130.684 | 0.7857 |
| 2 | feature selection | 130.5339 | 0.7857 |
| 2 | feature selection | 130.8196 | 0.7857 |
| 2 | feature selection | 131.1966 | 0.7714 |
| 2 | feature selection | 130.5681 | 0.7857 |
| 2 | feature selection | 131.0148 | 0.8286 |
| 2 | feature selection | 131.0194 | 0.8571 |
|  |  |  |  |  |

**III. Experiments ran on the dataset with other machine learning methods studied in the course and comparison of results with your focus machine learning method (at most 1 page)**

For all the testing below, the preprocessing contains the random permutation, which helps to improve the training the model quality, and also helps improve the accuracy a little bit.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ML Method**  **and Brief description**  **of the experiment** | **Pre-process** | **Eval**  **Covar metric 1 of 4** | **Eval**  **Covar metric 2 of 4** | **Eval**  **Covar metric 3 of 4** | **Eval**  **Covar metric 4 of 4** | **Time**  **taken** | **Observation on results** | **total**  **accuracy** |
| Decision Tree | two  classes | 0.2501 | 0.0440 | 0.0440 | 0.2493 | 31.233 s | Tree model is highly unbalanced | 0.5880 |
| Support Vector Machine | two  classes | 0.2501 | 0.0274 | 0.0274 | 0.2480 | 224.235 s | None | 0.5545 |
| K Nearest Neighbor | two  classes | 0.2501 | 0.0585 | 0.0585 | 0.2318 | 46.564 s | None | 0.6175 |
| Naive Bayesian Model | two  classes | 0.2501 | 0.0639 | 0.0639 | 0.2445 | 681.093 s | None | 0.6280 |
| Decision Tree | ten  classes | 7.9995 | 1.0019 | 1.0019 | 8.1481 | 32.854 s | Tree is mass up a lot | 0.2473 |
| Support Vector Machine | ten  classes | 8.2843 | 1.4301 | 1.4301 | 7.9618 | 1460.701 s | Takes too long time | 0.3045 |
| K Nearest Neighbor | ten  classes | 8.0883 | 1.5255 | 1.5255 | 6.8012 | 45.736 s | None | 0.2865 |
| Naive Bayesian Model | ten  classes | 8.4339 | 2.0857 | 2.0857 | 8.8248 | 697.823 s | None | 0.2895 |
|  |  |  |  |  |  |  |  |  |

**IV. Meta-Learning Experiments and Results: Combining Multiple Models (at most 1 page)**

[This section should contain a summary of the results of your meta-learning experiments and a thorough comparison with the results of your experiments in sections II and III above.]

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Meta-Learn. Method**  **and Brief description**  **of the experiment** | **Pre-process** | **Eval**  **Covar metric 1 of 4** | **Eval**  **Covar metric 2 of 4** | **Eval**  **Covar metric 3 of 4** | **Eval**  **Covar metric 4 of 4** | **Time**  **taken** | **Observation on results** | **total**  **accuracy** |
| Tree Bagging | two  classes | 0.2501 | 0.0806 | 0.0806 | 0.0806 | 6.349 s | not bad | 0.6605 |
| AbaBoost  M1 | two  classes | 0.2501 | 0.1042 | 0.1042 | 0.2498 | 530.234 s | very good | 0.7085 |
| Rubust  Boosting | two  classes | 0.2501 | 0.0973 | 0.0973 | 0.2501 | 526.225 s | good | 0.6945 |
| Stacking  KNN and  Naive Bayesian | two  classes | 0.2501 | 0.0758 | 0.0758 | 0.2468 | 531.049 s | better than each single prediction | 0.6520 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**V. Analysis of All of Your Results and Conclusions (at most 1 page)**

[1. Analyze the effect of varying parameters/experimental settings on the results.

2. Analyze the results from the point of view of the Dataset Domain, by describing patterns described by the machine learning models constructed.

3. Compare results you obtained from using different machine learning techniques, different pre-processing, different dimensionality reduction, and different post-processing (if applicable).

4. Analyze the effects of using meta-learning.

5. Include and explain (some of) the best / most interesting results you obtained in your experiments.

6. Include visualizations.