CSE-584 - MACHINE LEARNING

HOMEWORK - 1

Dhanush Gurram
PSU ID: 946870078
Email ID: dfg5539@psu.edu

❖ **Paper - 1**: *Adversarial Active Learning for Deep Networks: a Margin Based Approach*

➢ **What problem does this paper try to solve, i.e., its motivation**

Adversarial Active Learning for Deep Networks: A Margin-Based Approach gives an emphasis on the problem of learning how to train the deep neural networks, especially convolutional neural networks (CNNs) using a very less labelled input while maintaining or improving model accuracy. Standard active learning techniques find deep networks to be a challenging problem due to their complex architectures and very high-dimensional decision boundaries. The primary objective of this study is to develop a more efficient active learning method, by reducing the expense of data labelling and maximising the selection of beneficial samples, which would eventually enhance the overall performance of deep learning models.

➢ **How does it solve the problem?**

This paper gives an innovative active learning technique known as DeepFool Active Learning (DFAL), which estimates samples' proximity to a deep neural network's decision boundaries using the adversarial examples. DFAL uses the perturbations generated by adversarial attacks (produced by the DeepFool algorithm) to identify the unlabeled samples that are nearest to the decision border, as compared to perceiving adversarial examples as causing damage. The theory is about labelling these high-impact samples and the decision boundary can be improved more successfully, which leads to an overall enhancement in the model's performance even with fewer labelled cases.

Here, in this paper initial samples and their adversarial counterparts are added to the labelled dataset by DFAL, which focuses more on choosing samples that are most vulnerable to adversarial attacks. This pseudo-labeling method strengthens the network's efficiency while also fortifying its robustness, allowing it to generalise better to new data.

➢ **A list of novelties/contributions**

- Adversarial Examples in Active Learning: It illustrates the notion of directing active learning through the use of adversarial attacks (using the DeepFool algorithm). DFAL assists in locating

and choosing the most illuminating samples for labelling by providing an estimate of the distance to the decision boundary.

- Margin-Based Learning for Deep Nets: Based on margin theory, DFAL emphasises the significance of selecting samples that are close to the decision boundary in order to improve model performance. It's among the first systems to use convolutional neural networks (CNNs) with margin-based learning.
- Pseudo-Labeling with Adversarial cases: An innovative method that boosts the robustness and generalizability of the model by labelling adversarial cases alongside their original equivalents.
- No Hyperparameter Tuning Needed: Unlike many active learning methods, DFAL doesn't require hyperparameter tuning, making it more practical and easier to implement.
- Strong Empirical Results: DFAL has been rigorously tested on datasets like MNIST, Shoe-Bag, and Quick-Draw, where it outperforms leading techniques such as BALD, CORE-SET, and uncertainty sampling.

➢ **What do you think are the downsides of the work?**

Here are some potential drawbacks of the DFAL method:

- DFAL improves how data points are selected for labelling, but one of its challenges is that generating adversarial examples can be slow, especially when working with large datasets, complex networks, or multi-class scenarios. Since the method is mainly designed for convolutional neural networks (CNNs), it might not perform as well with other types of models, and its effectiveness beyond CNNs hasn't been thoroughly tested.
- While DFAL shows promising results in experiments, the paper doesn't provide strong theoretical evidence to explain why it consistently outperforms other methods. Most of its success is based on experimental findings. Another concern is that adding adversarial examples to the labelled dataset could cause the model to overfit, making it less capable of generalising to regular data variations.

In short, DFAL offers a fresh and effective approach to active learning for CNNs, but there are practical limitations, such as the time it takes to compute adversarial examples, concerns about generalising to other types of models, and the risk of overfitting.

❖ **Paper - 2**: *Active Learning For Convolutional Neural Networks: A Core-Set Approach*

➢ **What problem does this paper try to solve, i.e., its motivation**

This paper explores the challenge of reducing the need for massive amounts of labeled data when training deep convolutional neural networks (CNNs). Labelling data is not only costly but also time-consuming, yet deep learning models still rely heavily on large labelled datasets to perform well. The authors aim to find a more efficient way to select the most valuable data points for labelling, ultimately lowering the labelling cost while boosting model performance. Existing active learning methods for CNNs are often ineffective, especially in batch scenarios where multiple data points need to be labelled at once. This paper introduces a new approach to address these shortcomings.

➢ **How does it solve the problem?**

The authors suggest redefining active learning as an issue of core-set selection. Identifying a limited subset of data points in order to ensure a model trained on it performs similarly to a model trained on the whole dataset is known as the core-set technique. The discrepancy between the model's performance on the chosen subset and the entire dataset is known as the core-set loss, and they present a theoretical approach that aims to reduce it. This problem is mathematically equivalent to the k-Center problem, which the authors solve using a greedy algorithm to approximate the optimal solution. The method selects diverse, representative samples that ensure the model generalises well to the remaining unlabeled data.

➢ **A list of novelties/contributions**

- Core-Set Approach for Active Learning: This study effectively redefines active learning for deep CNNs as a core-set selection problem.
- Theoretical Bound on Core-Set Loss: Based on data geometry, this theoretical framework limits the core-set loss and determines the performance of the chosen subset.
- k-Center Problem in Active Learning: The core-set selection problem in active learning is solved through reducing it to the k-Center problem and approximating the answer with an effective greedy method.
- CNNs Using Batch Mode Active Learning: The method is more useful for large-scale applications since it is made to operate in batch settings, where multiple points are marked simultaneously.
- Empirical Validation: Demonstrates improved performance on multiple image classification datasets, including CIFAR-10, CIFAR-100, and SVHN, in comparison to current approaches.

➢ **What do you think are the downsides of the work?**

Here are the few drawbacks of the paper:

1. The approach requires resolving the k-Center combinatorial optimization problem, which can be computationally demanding, especially when dealing with large datasets.
2. The theoretical analysis makes the assumption that there is zero training error, which may not always be the case in practical situations and therefore restrict the method's application.
3. Although the strategy is effective for CNNs, its ability to generalise to other models or applications outside image classification is uncertain.
4. The k-Center problem's greedy algorithm gives a 2-OPT solution, which is suboptimal in comparison to the actual optimal solution and may include performance trade-offs.

❖ **Paper - 3**: *Support Vector Machine Active Learning with Applications to Text Classification*

➢ **What problem does this paper try to solve, i.e., its motivation**

The paper aims to address the challenge of reducing the dependency on large amounts of labelled data in supervised learning tasks, particularly for Support Vector Machines (SVMs). Labelling data is expensive and time-consuming too, yet many traditional machine learning models often require extensive labelled datasets to perform well. The authors focus on optimising this process by exploring active learning, where the model intelligently selects the most informative data points to label, rather than using a random set. This approach is of high

importance in text classification tasks where large amounts of unlabeled data exist, but manually labelling them is impractical.

➢ **How does it solve the problem?**

The paper solves the problem by teaching Support Vector Machines (SVMs) to be smarter about which data to learn from. Instead of randomly picking data to label, the model carefully selects the most helpful data points that will improve its understanding the fastest. It uses something called version space—basically, all the possible models that could be right based on what it's already seen. The model's job is to narrow this down as efficiently as possible. To do that, it uses three different methods (Simple Margin, MaxMin Margin, and Ratio Margin) to decide which data point to label next, focusing on the ones that will make the biggest impact on learning. This approach means the model can get really accurate with far fewer labelled examples, saving a ton of time and effort, especially for tasks like sorting through text.

➢ **A list of novelties/contributions**

New Active Learning Algorithm for SVMs: The authors created a fresh approach to active learning specifically designed for Support Vector Machines, making the process of picking which data points to label smarter and more efficient.

- Version Space Concept: The algorithm is based on a cool idea called version space, which helps the model figure out the most useful data points for learning by focusing on the area where the possible models lie.
- Three Active Learning Strategies: The authors have proposed three different ways the model can choose data points:
    1. Simple Margin: Pick the data point closest to the decision boundary.
    2. MaxMin Margin: Balance the remaining uncertainty after each label.
    3. Ratio Margin: Keep the reduction of uncertainty as balanced as possible.
- Real-world Impact: Their methods were tested on real-world text classification tasks, like the Reuters-21578 dataset and Newsgroups, proving that they could dramatically reduce the number of labelled examples needed.
- Broad Applicability: The algorithm works for both inductive learning (making predictions on new data) and transductive learning (labelling data within a specific dataset), showing its versatility.
- Significant Reduction in Labelling: Their experiments demonstrated that active learning could massively cut down the number of labelled data points required, making the learning process much more efficient.

➢ **What do you think are the downsides of the work?**

1. Computational Complexity: Although the Ratio Margin and MaxMin approaches are more precise, they have a drawback in that they need a significant amount of processing. They need to train several SVMs for each query, which becomes a significant issue as the quantity of unlabeled data points increases. When dealing with bigger datasets, this causes them to be slower and less scalable.
2. Simple Margin Instability: Although the Simple Margin approach is quicker, it has drawbacks of its own. In some circumstances, it might be unreliable and result in subpar performance. Its inconsistent performance was evident when compared to the other approaches in certain Newsgroups datasets.
3. Assumption of Linear Separability: The method operates on the presumption that a straight line (in the feature space) may be used to cleanly separate the data. Unfortunately, in more

complicated real-world settings, that doesn't always happen, which reduces the flexibility of the strategy.

4. Limited Exploration: Because Simple Margin likes to be safe, it doesn't go as far as it might in exploring the data space. In the long term, this lack of vigorous research becomes less successful since it might occasionally result in poor decisions about which data points to categorize.