Effective Writing in ML: Myths and Misconceptions

Exercise 1

Each question is one of the following types: Simple Counting, Complex Counting, Condition of Road, Condition of Entire Image and Yes/No type of question. **<u>The dataset distribution is shown in Figure 2. / Figure 2 shows the dataset distribution.</u>** These questions result in 56 probable labels, out of which the first five are represent non-numerical answers to the questions: “flooded”, “non-flooded”, “flooded,non-flooded”, “Yes” and “No”.

Exercise 2

<u>There</u> has been an extensive use of Unmanned Aerial Vehicles in search and rescue missions to distribute first aid kits and food packets. **<u>It</u>** is important that these UAVs are able to identify and distinguish the markers from one another for effective distribution.

[unnumbered example]

With the help of this interpretation, a natural better **<u>approximation for PIDCs can be obtained</u>** by adding an integral term to FedCostWAvg.

With the help of this interpretation, **<u>we can approximate PIDCs</u>** better and more naturally by adding an integral term to FedCostWAvg.

Exercise 3

Taking inspiration from these previous works, our algorithm **<u>learns</u>** the best-fitting parameters of a spectral filter on a corrupted dataset without supervision.

Taking inspiration from these previous works, the best-fitting parameters of a spectral filter **<u>are learnt</u>** on a corrupted dataset without supervision.

Exercise 4

There has been an extensive use of Unmanned Aerial Vehicles in search and rescue missions to distribute first aid kits and food packets. It is important that these UAVs are able to identify and distinguish the markers from one another for effective distribution. One of the common ways to mark the locations is via the use of characters superimposed on shapes of various colors which gives rise to wide variety of markers based on combination of different shapes, characters, and their respective colors. In this paper, we propose an object detection and classification pipeline which prevents false positives and minimizes misclassification of alphanumeric characters and shapes in aerial images.

Exercise 5

Machine Learning models are prone to fail when test data are different from training data, a situation often encountered in real applications known as distribution shift. While still valid, the training-time knowledge becomes less effective, requiring a test-time adaptation to maintain high performance. Following approaches that assume batch-norm layer and use their statistics for adaptation (Nado et al., 2020), we propose a Test-Time Adaptation with Principal Component Analysis (TTAwPCA), which presumes a fitted PCA and adapts at test time a spectral filter based on the singular values of the PCA for robustness to corruptions. TTAwPCA combines three components: the output of a given layer is decomposed using a Principal Component Analysis (PCA), filtered by a penalization of its singular values, and reconstructed with the PCA inverse transform. This generic enhancement adds fewer parameters than current methods (Mummadi et al., 2021; Sun et al., 2020;Wang et al., 2021). Experiments on CIFAR-10-C and CIFAR-100-C (Hendrycks & Dietterich, 2019) demonstrate the effectiveness and limits of our method using a unique filter of 2000 parameters.

Exercise 6

Object Detection in Aerial images is more challenging and demands separate attention because targets are small and sparse with concentrations over minority regions. Semantic segmentation and detection is integrated in [1] in order to improve performance. Paper [3] investigates misalignment between ROI and objects in aerial image detection. It introduces a ROI transformer to address this issue. A scale adaptive proposal network is also proposed for object detection in aerial images in [4].

Exercise 7

Test-time adaptation indicates methods tackling the domain gap during inference. TTT (Sun et al., 2020) augments the supervised training objective with a self-supervised loss using source data. Only the self-supervised loss keeps adapting at test time on target domain. It relies on predicting the rotation of inputs, a visual proxy task, but designing suitable proxy tasks can be challenging. Training parameters are altered during training and test-time adaptation. Test-time batch normalization (Schneider et al., 2020; Nado et al., 2020) allows statistics of batch norm layers to be tracked during the distribution shift at test time. TENT (Wang et al., 2021) exhibits entropy minimization at test time on feature modulators extracted from spatial batch normalization to adapt to distribution shift. Entropy minimization is a generic and standard loss for domain adaptation to penalize classes overlap. Information maximization (Krause et al., 2010; Shi & Sha, 2012; Hu et al., 2017) used by (Liang et al., 2020; Mummadi et al., 2021) involves entropy minimization and diversity regularization. The diversity regularizer averts collapsed solutions of entropy minimization. SLR+IT (Mummadi et al., 2021) argues that Information maximization compensates for the vanishing gradient issues of entropy minimization for high confidence predictions. Moreover, an additional trainable network shares the input samples with the tested network to partially correct the domain shift. Principal Component Analysis cuts out noisy eigenvalues to remove uncorrelated noise (Li, 2018; Murali et al., 2012). In addition, we propose to add fully test-time learnable parameters to reduce the remaining noise of corrupted data onto the spectral basis.

Exercise 8

The MSA encoder extracts image features. An FS method (or multiple) selects the best features. A classifier is trained on the best features. This is then used for prediction.

Exercise 9

(1) Extending the interpretation to the weighted averaging approach, it is natural to interpret the aggregation procedure introduced in (Mächler et al., 2021) as an approximation of a PID controller where the cardinality-dependent and the loss-dependent terms are respectively functioning as the proportional component and the derivative component in a PID controller (PIDC).

(2) Following approaches that assume a batch-norm layer and use their statistics for adaptation (Nado et al., 2020), we propose a Test-Time Adaptation with Principal Component Analysis (TTAwPCA), which presumes a fitted PCA and adapts at test time a spectral filter based on the singular values of the PCA for robustness to corruptions.

Exercise 10

Since we [use] traditional image processing for ROI Detection and unsupervised learning for segmentation, there [be] a possibility for false positives in our images. Reducing False Positives [be] important because it [be] very costly if the drone [keep] on arriving at the wrong locations.

We [do] false positive removal in two stages:

We [count] the number of regions present in the mask.