

BITS F464 - Machine Learning
I Semester 2021-2022
Assignment #2
Weightage: 15%
Due Data: 08th December, 2021

Active Learning (Implementation Assignment – 10%)

- (a) Convert a supervised learning problem (multiclass classification – having more than 3 classes) into an Active Learning problem by randomly removing class labels of data points (retain only 10% of labelled points). The removed labelled points will work as a human oracle!!
- (b) Use pool-based scenarios to label additional 10%, 20%, 30%, 40% data points.
 - i. use uncertainty sampling (Least confident, Smallest Margin Sampling, Largest margin sampling, & Entropy) to label points. Compare the three measures of informativeness.
 - ii. Use QBC (Vote Entropy & KL divergence) with at least 5 committee members, to label points. Compare the two measures of disagreement.
 - iii. What is the size (number of points) of the version space? Order points to label in such a way that the version space gets reduced by maximum with each point chosen to be labelled (greedy approach)
 - iv. Incorporate the additional labelled points (separately, the best from i & ii) into your model and compare with corresponding models trained with randomly chosen labelled points.
 - v. From 90% of unlabelled points, randomly pick 40% of the points and use clustering (using K-means with K=number of class labels). In each cluster, randomly label 20% of the points to label remaining points in the cluster. How accurate is the cluster-based labelling? How much saving it results in if each label costs you Rs. 100 and each labelling takes one hour.
- (c) How can we use stream-based active learning in the above scenario?

Bias in Machine Learning Systems (Reading Assignment - 5%)

Any kind of bias in ML systems is seen as a negative trait which is undesirable, and researchers have been working on defining, identifying, and mitigating bias [1-6]. ImageNet Roulette exposed the systemic bias in ML systems [4] and consequently, 600,000 images were removed from its “person” category [6]. Designing fair (bias free) ML algorithms is attracting lot of interest lately [7,8].

Understand the concept of Bias by reading the reference papers given below. You need to focus on different types of biases, problems due to presence of bias, identification of bias, and techniques to mitigating bias.

References for Bias in ML Systems

- [1] Baeza-Yates, R. Bias on the Web. Commun. ACM 61, 6 (June 2018), 54–61.
- [2] Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., Rieke, A. Discrimination through optimization: how Facebook’s ad delivery can lead to biased outcomes. In Proceedings of the ACM on Human- Computer Interaction 3 (2019); <https://dl.acm.org/doi/10.1145/3359301>.
- [3] Buolamwini, J., Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of Machine Learning Research 81 (2018), 1–15; <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.
- [4] Crawford, K., Paglen, T. Excavating AI: The politics of images in machine learning training sets. The AI Now Institute, New York University, 2019; <https://www.excavating.ai>.
- [5] Raji, I., Buolamwini, J. Actionable auditing: investigating the impact of publicly naming biased performance results of commercial AI products. In Proceedings of the 2019 AAAI/ACM Conf. AI, Ethics, and Society, 429–435; <https://dl.acm.org/doi/10.1145/3306618.3314244>.
- [6] Small, Z. 600,000 images removed from AI database after art project exposes racist bias. Hyperallergic, 2019; <https://hyperallergic.com/518822/600000-imagesremoved-from-ai-database-after-art-projectexposesracist-bias/>.
- [7] Barocas, S., Hardt, M., Narayanan, A. Fairness and machine learning: limitations and opportunities, 2019; <https://fairmlbook.org>.
- [8] Bellamy, R.K.E. et al. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. 2018, arXiv; <https://arxiv.org/abs/1810.01943>.

What you need to submit:

1. Details of the classification problem chosen including the metadata about the dataset.
2. A report on active learning assignment containing all details of how you worked on the assignment. Include all results in the report.
3. A report on Bias in ML systems

Navneet Goyal