ELV832: Assignment, Part 1 Report

Name: Lokesh Varadaraj
Entry Number: 2022PH11858

Data Generation:

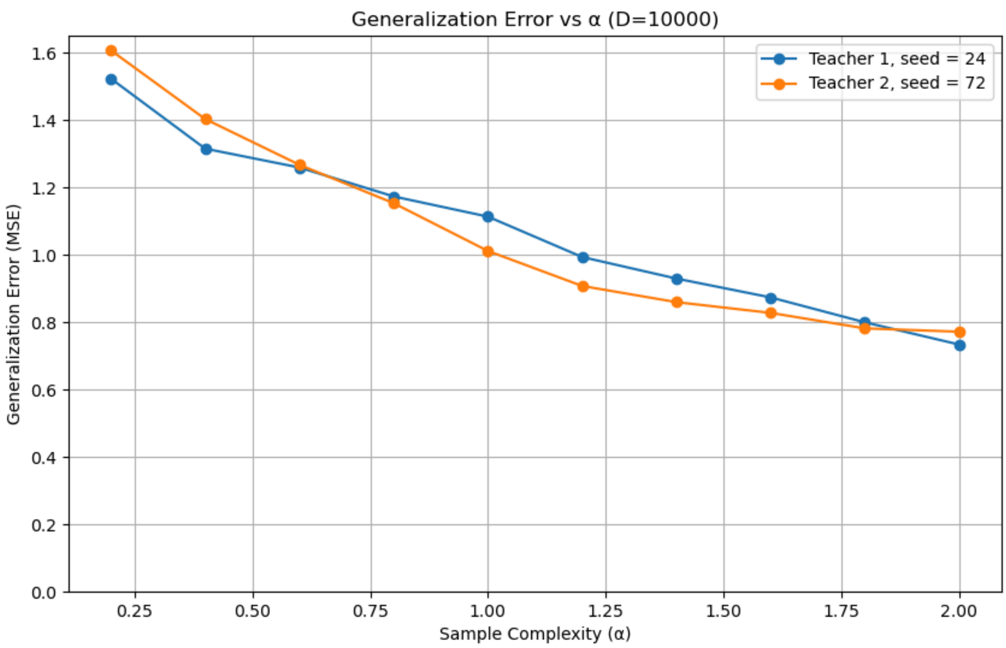
- We generate a random vector of binary weights, i.e., +1/-1 with dimensionality D, This constitutes our *teacher* model (binary perceptron) .
- We generate a training set of size 2D plus some test data of size 0.2D. Hence, generate an overall data set of size 2.2D.
- All feature values should are sampled IID from a standard normal distribution
- Each feature vector is passed through the teacher model (with sign() activation for the binary perceptron) to obtain its corresponding class label.

Code:

- Attached are two python files, Assignment1.py and Assignment1_opt.py.
- Assignment1.py is a naive implementation and Assignment1_opt.py is an optimised implementation. The attached plots are from Assignment1_opt.py.

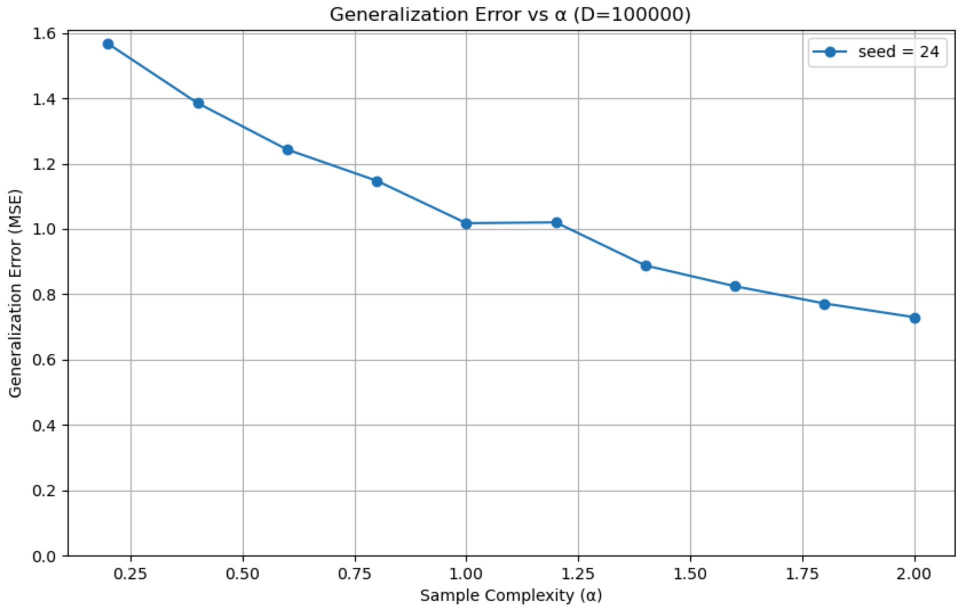
D = 1e4 for two different teacher models:

- Below is a plot of generalization error(MSE) vs sample complexity for D = 1e4.
- The trend followed is the same for both teacher models (with different random seeds)
- The trend followed is similar to that of fig.2 of Barbier et al where we observe a decrease in Generalization error with increasing sample complexity. However, we do not observe a phase transition for D = 1e4 when using logistic regression.



D = 1e5 for data generated from a teacher model:

- The trend is similar to that of D=1e4, we do not observe a phase transition here.
- Scaling to even larger D is challenging due to computational constraints, particularly memory limitations.



Why do we not observe a phase transition:

- For the above setting, we theoretically expect a phase transtion ([Barbier et al]) at some alpha. However, we do not observe one for both D = 1e4 and D = 1e5.
- Using the Bayes optimal algorithm, we observe a phase transition at alpha = 1.249 which is the smallest possible alpha to observe a phase transition information theoretically.
- Using GAMP, we also observe a phase transition at a slightly higher alpha =1.493, ie we observe perfect generalisation only above alpha = 1.493 using GAMP.
- Logistic regression does not exhibit a phase transition in this setting, possibly due to differences in weight constraints between the teacher model and logistic regression.
- In the teacher model, the weight vector W is binary (+1/-1), and the class label is given by y = sign(W · X), where X is a (2.2D, D) matrix with i.i.d. Gaussian entries.
- Logistic regression learns a weight vector W_LR and bias b, predicting y_prob = sigmoid(W_LR · X + b) and classifying based on y_pred = sign(W_LR · X + b). Since W_LR and b take real values, the model may fail to perfectly recover the binary W, leading to limited generalization.
- This issue is better handled in GAMP, where W is explicitly constrained to be binary.
- The paper assumes D -> inf limit for all it's theoretical findings. Trying a larger value of D might help observe a phase transition even using logistic regression, but this is doubtful as even for D = 1e5 no such transition was observed.

Ideas to change the settings of the experiment to observe a phase transition:

- Incorporate knowledge of the teacher model when designing the student model, ensuring a well-chosen prior.
- Try training for even larger D if computationally feasible.