Homework 2 -Statistical Learning Theory - Spring 2025

Essay Question: *VC theory has had mixed results in successfully explaining the generalization ability of various neural network model sets..* Write a one paragraph essay that starts with this sentence and goes on to identify and describe the successes and failures of VC theory in predicting the generalization ability of various neural network models. Be sure to provide examples supporting your assertions and conclude your essay with a single sentence summarizing the main findings of your essay.

Problem 1: Consider a model set $\mathcal{H}$ with a finite VC dimension of 3 and let $h^* \in \mathcal{H}$ be a model that minimizes the empirical risk on a training dataset, $\mathcal{D}$, with 10000 samples.

1. Determine the generalization error, ϵ, for this model, h^*, with a 90% confidence level.
2. Let $\mathcal{D}'$ be a dataset with N' samples that is statistically independent of the original dataset, $\mathcal{D}$, used to obtain the model h^*. Determine the size of the dataset, $\mathcal{D}'$, needed to obtain the generalization error that is 10 times smaller than the generalization error you found in the preceding problem with 90% confidence.
3. How many times larger or smaller is N' than the size of the original dataset $\mathcal{D}$? Comment on the significance of this difference?

Problem 2: A coin is weighted so its probability of landing heads is 20%, independently of the other flips. Suppose a coin is flipped 15 times. Determine the probability that 75% of the 15 tosses are heads. Then use the Markov, Chebyshev, and Chernoff inequalities to bound the probability that 75% of the 15 tosses are heads. Compute the ratio of each bound over the actual probability of the event and comment on how much better the Chernoff bound is versus the other two.

Hint: To compute determine the tightest Chernoff bound you will need to know that the moment generating function of a Binomial random variable $X \sim \text{Binom}(n, p)$ is $\mathbb{E}[e^{\lambda X}] = ((1-p)+pe^{\lambda})^n$ where $\lambda > 0$.

Notebook Assignment 1 (TensorFlow): This assignment uses the fashion-mnist dataset. Fashion-MNIST is a dataset of Zalando's article images; consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28 by 28 grayscale image associated with a label from 10 classes. The specific tasks to be completed for this exercise are enumerated below.

1. **TO DO:** Load the Fashion-MNIST dataset's training and testing data (samples and targets). Retype the input samples as float32 and check the shape of the arrays for the training/testing samples and targets.
2. **TO DO:** Find an image for each type of garment and display that image with the title of that garment over the top of the image. Arrange your 10 images in a 2 (rows) by 5 (columns) grid.
3. **TO DO:** Instantiate a tensorflow model with two dense layers. The input layer has shape (28*28,). The next dense layer has 128 nodes with an relu activation. The last (output) layer

has 10 nodes with a softmax activation. Comple the model with an adam optimizer, sparse categorical crossentropy loss function, using ”accuracy” as the performance metric. Build the model and print the model summary.

4. **TO DO:** Reshape the training and test samples to have shape (60000,28*28). Rescale the sample to [0,1] and make sure they are typed as float32. Split the training data into a p-training and validation set using a 2/3 split. Create a callback to save the ”best model” with the lowest validation loss. Then Train the model on the p-training data using the validation data to create the validation curves. Train your model for 30 epochs assuming a batch size of 512 and return the history object.

5. **TO DO:** Create a function `plot_training_curves` that takes the `history` object and plots the ptraining/validation loss in one figure and the ptraining/validation accuracy in the other figure. Then evaluate the best model's accuracy on the test data.