***ECE517 Fall 2020***

***Homework 3.1***

***Robert M. Taylor***

The **risk** or expected risk is the amount of loss/risk we expect for the data in a particular distribution (P). For instance, the expected risk R(f, P) is calculated by looking at the positive loss function to measure how well it is aligned with the actual data.

**Emperical risk**, on the other hand, is the mean loss for the data. This is simply the summation of the individual loss estimators.

***Eq. 1***

The **complexity** of a model refers to the number of parameters that the given model is given during the learning process. It can be considered as the dimension of space that the model can take. If a model is not complex enough, the model will not correctly learn the labels and will perform poorly. On the other hand, if a model is too complex, it will try to fit the data to parameters that are not actually significant. Although this model may perform well on the training set, it will perform poorly on the test set. This is because the model was too complex, had too many parameters, and thus the model used insignificant parameters to form the model and will **overfit** the data. Meaning the model has is learning examples, rather than unearthing features pertinent to the proper classification.

If we choose a confidence interval (n) then the above empirical risk (**eq. 1**) can be rewritten as

***Eq. 2***

In equation 2, the h-term is called the Vapnik Chervonenkis (VC) dimension. This **VC dimension** is the value of the complexity of the model described above. It is the maximum number of points that can be shattered (labelled to a specific class) by the model. These points of training data can be shattered by a hyperplane that separates the 2 classes of data. The hyperplane/s in the dimension of space follow the below theorem and corollary:

***Theorem 1:*** Consider some set of m points in . Choose any one of the points as origin. Then the m points can be shattered by oriented hyperplanes if and only if the position vectors of the remaining points are linearly independent. Reference!!!!

***Corollary:*** the VC dimension of the set of oriented hyperplanes in is n+1, since we can always choose n + 1 points, and then choose one of the points as origin, such that the position vectors of the remaining n points are linearly independent, but can never choose N + 2 such points (since no N+ 1 vectors in can be linearly independent) Reference!!!!

Since the hyperplane divides the two classes of data, we wish to optimize or maximize the margins between the two closest and different data points, so the hyperplane is the furthest possible distance from both classes.

INTRODUCE SVM CRITERIA

**Support Vector Machines (SVM)s** rely on the ability to minimize the empirical risk and maximize the margins of the hyperplane/s.

DEVELOP ANALYSIS THAT LEADS TO THE DUAL SOLUTION OF THE SVM

DESCRIBE THE PROPERTIES OF THE SUPPORT VECTORS