**ECE 395**

**Assignment 3.1**

**Author: Ralph Mora**

**Date: October 2nd, 2023**

**Risk and Empirical Risk.**

In Machine learning, Risk refers to the expected loss or error a machine might encounter given unseen/real world data. Empirical risk is calculated by processing a set of training data that will allow us to get an estimate of the true risk of the machine. Although an Empirical risk equal to zero would be ideal, the complexity of the machine can increase to the point where the computer is unable to process the calculation in a moderate amount of time.

**Complexity and overfitting.**

Empirical risk and Complexity are inversely related. As the Empirical risk approaches zero, we begin overfitting our data. When we overfit our data we can overcompensate and miss many underlying patterns. Finding these underlying patterns can help us optimize or “fit” our machine so it can process a data set in the shortest amount of time with the least error.

**Occam’s Razor and the Vapnik-Chervonenkis (VC) dimension**

Occam’s razor is also known as the law of succinctness, it states that we should chose the simplest methods that adequately explain our observed data as opposed to taking the most complex. This is an important concept in machine learning because overfitting large data sets can lead to computations that could take years to complete even with modern day computers.

This is important to the Vapnik-Chervonenkis (VC) dimension which is a way of characterizing the risk of overfitting. In a space of dimension N, any set of N + 1 points can be perfectly classified.

This is important to the Vapnik-Chervonenkis (VC) dimension which is a way of characterizing the risk of overfitting. In a space of dimension N, any set of N + 1 points can be perfectly classified.

**Vapnik-Chervonenkis** (**VC) Theorem**

The VC Theorem tells us that the true risk of a classifying machine can be bounded by the empirical risk plus a term that depends on the VC dimension which we call the structural risk.

A screenshot of a computer

Description automatically generated

Where the far-right term is the structural risk, h is the machines complexity, and N is the number of samples. We can see that the structural risk is proportional to the machines complexity and is inversely proportional to the number of data samples and as we discussed earlier, the empirical risk has the opposite relation. So, minimizing the true risk requires finding the best balance between the empirical risk and structural risk.

**Support Vector Machines**

The Support Vector Machine (SVM) consists of minimizing the empirical risk and structural risk through margin maximization.

Minimize: Lp(w, €n) = ||w||2 + n

Subject to: yn(wT xn + b) > 1 − €n and €n ≥ 0

Where C is a free tradeoff parameter. We then perform Lagrange multipliers to change the constrained problem into an unconstrained one.

Lp(w, €n, αn, µn) = ||w||2 + €n - αn(yn(wTxn + b) -1 + €n) - µn€n

1. Δw Lp(w, €n, αn, µn) = w - αnynxn = 0
2. Δ€n = C - αn - µn = 0
3. Δb = αn yn = 0
4. µn€n = 0
5. αn(yn(wTxn + b) -1 + €n) = 0
6. αn ≥ 0, µn ≥ 0, €n ≥ 0

we can then plug the equation w = αnynxn into our estimator yk = wTxk+ b.

yk= αnynxnTxk+ b = αTYXTxk + k

Lp(w, €n, αn, µn) = ||w||2 + €n - αn(yn(wTxn + b) -1 + €n) - µn€n

We can then plug yk into our Lagrangian and reduces to the following dual solution, which ensures existence and uniqueness because the matrix αTYXTXYα > 0

Ld = αTYXTXYα+αT1

**Support Vectors**

Support vectors are pivotal data points within an SVM, exerting a significant impact on determining the location and direction of the decision boundary. Their selection is based on their location as the most demanding points to accurately classify or predict,