ECE 517 HW 3.1

Joseph Hilland

November 2023

1 Assignment Outline

This assignment is for the student to summarize the theoretical part of this module in a format that can be later inserted in an article (Word, LaTex...). This assignment, then, will be used in next module as part of the corresponding assignment.

There is no rubric for this assignment, and the only objective is to summarize the theory using a logic structure from the criterion to the implemented SVM. In next module, the assignment will be corrected to fit all the expectations of the theoretical part of a journal paper, and assessed following the corresponding rubric.

Outline of the assignment

Summarize the theory of this module in a maximum of three pages using the following structure:

1) Explain the concepts of Risk and Empirical Risk.

2) Explain the concepts of complexity and overfitting.

3) Introduce the concept of VC dimension.

4) Enunciate and interpret the VC theorem that describes the bound on the actual risk.

5) Introduce the SVM criterion.

6) Develop the analysis that leads to the dual solution of the SVM, and its main results.

7) Describe the properties of the Support Vectors.

1.1

Risk and Empirical Risk Risk and empirical risk are two methods to measure the precision and accuracy of a particular algorithm. There are typically three different sets of risks within machine learning. This includes, actual, structural and empirical.

Empirical risk can also be known as the training or sample risk. The model being trained can be measures for its average error over the dataset being used. This can be summarized as how well a model fits the training data being used.

Actual risk, is the measure of the models performance on all inputs from the entire population. It can be approximated with help from the structural and empirical risk.

Using these risk techniques we can identify how well a model will perform and if the algorithm is underfit or overfit for our data.

1.2 Complexity and Overfitting

complexity refers to the sophistication of a model. Overfitting can occur where a model learns the training data too well. This ends up capturing noise or random fluctuations within the training data. This can result in a model that performs well on training data, however it will fail when given new or unseen data. This phenomenon occurs when a models algorithm becomes too complex. Keep the model as simple as possible.

1.3 Vapnik–Chervonenkis (VC) Dimension

The VC dimension was developed during the 1960s-1990s. It was developed by Vladimir Vapnik and Alexey Chervonenkis. It was meant to be a statistical approach to classification of problems.

The VC dimension can determine the max number of vectors that can be shattered by a hyperplane and gives a measure of the complexity of linear functions. If the VC Dimension of an estimator is higher than the number of vectors to be classified, then the estimator is guaranteed to overfit.

1.4 VC Theorum

With the Vapnik–Chervonenkis theorem, we need to define the linear empirical risk as:

$$R_{emp}(\boldsymbol{\alpha}) = \frac{1}{2N} \sum_{n=1}^{N} \left| \; y - f(\mathbf{x}, \boldsymbol{\alpha}) \; \right| \tag{1}$$

where $f(\cdot)$ is defined so that the loss function $\left| \; y - f(\mathbf{x}, \boldsymbol{\alpha}) \; \right|$ can only take the values of 0 or 1. Then, with the probability of $1 - \eta$, the following bound holds:

$$R(\boldsymbol{\alpha}) \leq R_{emp}(\boldsymbol{\alpha}) + \sqrt{\frac{h(\log(2N/h) + 1) - \log(\eta/4)}{N}} \tag{2}$$

This is a bound on the risk with probability $1 - \eta$, so it is therefore neither guaranteed nor dependent on the probability distribution. While the left side is not computable, the right one can be easily computed provided the knowledge of h, where the second term of the right side is called the structural risk (R_s). The inductive *Principle of Structural Risk Minimization* consists then on choosing a machine whose dimension h is sufficiently small, so that the bound on the risk is minimized.

1.5 SVM Criterion

A support vector machine constructs a hyperplane or possibly a set of hyperplanes. These hyperplanes can be used for regression or classification. The optimal hyperplane requries the usage of two other hyperplanes that are parallel fro it on either side. These other two hyperplanes used for support, are located within the most extreme points between the classes.

1.6 SVM Solution, Results

For the dual solution, the emprical risk and structural risk needs to be minimized. This is done through te margin maximization which can be seen below:

$$\text{minimize } L_p(\mathbf{w}, \xi_n) = \frac{1}{2}\|\mathbf{w}\|^2 + \ C\sum_{n=1}^{N} \xi_n \tag{3}$$

$$\text{subject to } \begin{cases} y_n\left(\mathbf{w}^\top \mathbf{x}_n + b\right) > 1 - \xi_n \\ \xi_n \geq 0 \end{cases}$$

C is a free parameter, p denotes the primal and En is the distance where a datapoint exceeds the support hyperplane towards the other class.

Solving for w will result in the Karush-Kuhn-Tucker condition that creates a point over the domain of the criteria.

1.7 SV Properties

Support vectors are datapoints near hyperplanes that satisfy certain conditions. This is seen in support-vector machines. If a datapoint is well within the boundary (support hyperplane), the penalizing factor ξ_n is 0. Otherwise, if the datapoint is on the other side, this factor ξ_n is equal to its distance between the datapoint and the support hyperplane, i.e., $\xi_n \geq 0$ and $\alpha_n = C$. If a sample is on the margin, $0 < \alpha_n = C$. Finally, if a sample is outside the margin, $\xi_n = 0$ and $\alpha_n = 0$.