ELL888 Indian Institute of Technology Delhi
MCM Kernel Optimization
Scribed by: *Mridul Gupta (2021AIZ8322)*
Instructors: Sandeep Kumar and Jayadeva

Contents

1 Structure

Structure these points later

☒ Class separability might be worse in the feature space.

- Start with we want to get a bound on the risk on the test data. This requires knowing the distribution from which the data is being generated.
- We cannot know the distribution parameters. So, we estimate the actual risk using empirical risk which is the risk given some (training) samples from the distribution.
- Risk bound equation in VC slt.pdf
- Highlight two parts, the empirical risk and the structural risk
- make note that to reduce risk, we must reduce h (VC-Dimension)
- Mention VC-Dimension bound for Δ-margin classifiers.
- Mention that in general it will be $n + 1$ but it can be bound by reducing $\frac{R^2}{\Delta^2}$
- Show the simplified graph
- This transformation might not take to same dimensional space, might be higher dimensional. This function is a Kernel, that needs to be optimized to optimize structural risk.

- Also mention examples why sometimes even an infinite dimensional kernel might not work. And why we need Kernel Optimization.
- Go into the details of how to do this.
- Mention appropriate theorems, definitions and lemmas in between.
- Add a key takeaways section.

Refined list to structure

- A little intro to MCM (slt, minimize vc dim bound, structural risk minimization).
- What is kernel optimization, why it is needed
- Amari's idea, magnifying around support vectors
- Xiong's idea of using a scatter matrix, Fisher discriminant
- Xiong's idea of creating a kernel scatter matrix, how it's equal to normal scatter
- Empirical feature space, matrix equations
- solving generalized eigenvalue problems
- Getting the MCM with optimized kernel

2 Introduction

Support Vector Machines (SVMs) are older than the state of Haryana (the linear variant at least), and the kernel version of SVMs were proposed by Vapnik *et al.*in 1992. Vapnik and Chernvonenkis along with others also developed theory key thoeretical concepts that constitute statistical learning theory. In what follows, we will discuss the theoretical and practical advancements since then that led to the work titled "Kernel optimization using conformal maps for the minimal complexity machine" by Jayadeva *et al.*.

Since this is the congruence of two paths, one leading to MCMs and the other leading to kernel optimization in a data dependent way, in the following we first discuss MCMs, and then kernel optimization, and finally how the two fit together.

3 Where does MCM come from?

The key concepts that lead to MCMs were available in statistical learning theory long ago, but weren't applied until 2015. So let's discuss the key concepts that naturally lead to minimal complexity machines. [1]

4 3

References

[1] Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to algorithms.*. MIT press. 2009 Jul 31