# Generalization Ability of Majority Vote Point classifiers

Akshat Agarwal    Rahul K Sevakula

Department of Electrical Engineering
Indian Institute of Technology Kanpur

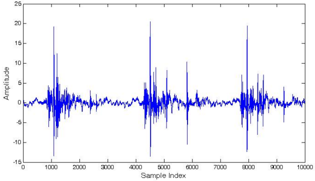
August 22, 2015

# Outline

# Health Monitoring of Industrial Machines

- Health monitoring of machines has been a subject of great interest for many decades.
- Diagnosis or prognosis of machine components is generally done by analysing various machine parameters like vibration, acoustics, temperature, pressure etc.
- A common observation with these machine parameters is that they can be very inconsistent.
- An example of this inconsistency can be seen in acoustic fault diagnosis of air compressors, where the nature of acoustic recordings changes with time, wear and tear of the machine, and even on its repair.

# Need for Classifiers with High Generalization



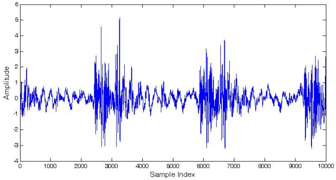




Figure : Though both recordings are taken in the same machine state and from the same sensor position, they are quite different from each other

- In such a situation, performing real time diagnosis with low level features [1] can be very difficult. This brings out the need for a classifier that is highly generalized.
- Classification problems with small number of samples and high dimensionality also need highly generalized classifiers.

# Vapnik-Chervonenkis (VC) Dimension

- A measure of the capacity or complexity of a classification algorithm
- It is defined as the cardinality of the largest set of points that the algorithm can shatter [2].
- A set of points is said to be shattered by a class of functions if a member of the class can perfectly separate them no matter how we assign a binary label to each point.

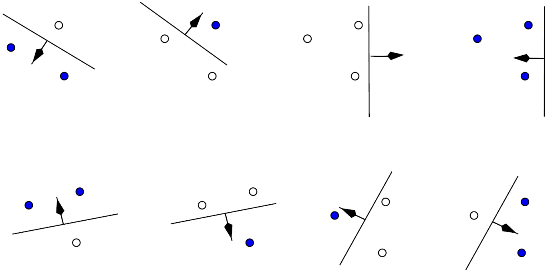


Figure : VC dimension of a 2D space is 3 [3]

# Growth function

The growth function $\Pi_H(m)$ of a classifier space $H$ is the maximum number of ways into which $m$ points can be classified by $H$.

$$\Pi_H(m) = max\{|\Pi_H(S)| : S \subseteq \Omega, |S| = m\} \tag{1}$$

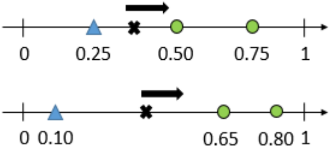
where $\Omega$ is the sample space $\{0, 1\}^m$.
Therefore, if $VCD(H) = d$, when $m \leq d$, $\Pi_H(m) = 2^m$. When $m \geq d$, an upper bound can be applied on the growth function using Sauer's lemma [4],

## Sauer's lemma

$$\Pi_H(m) \leq \Phi_d(m) := \sum_{i=0}^{d} \binom{m}{i} \leq \left(\frac{em}{d}\right)^d \tag{2}$$

# The Majority Vote Point (MVP) classifier

- Domain of the hypothesis space is in $\mathbb{R}$.
- The range for class labels is $\{0, 1\}$, meaning there are only two classes spanning the entire data, namely 0 and 1.
- The classifier will be trained on a single feature in the data and minimization of training error will be its objective.
- Learning the individual classifiers is similar to finding a threshold point on a line that has direction information regarding the class label.
- The number of classifiers selected for majority voting will be equal to the number of features in the data.


0
0.25
0.50
0.75
1
0
0.10
0.65
0.80
1

# Upper Bound on VC dimension

Consider a hypothesis space $H$ with $VC(H) = d$ and let $H_N$ be a majority vote classifier combining N ($\geq 1$) classifiers in $H$.
Let $VC(H_N) = p$. Then there exists a subset $S$ of the sample space $\Omega$ with $p$ elements such that $S$ is shattered by $H_N$. Then, from (1)

$$\Pi_{H_N}(p) = 2^p \tag{3}$$

Since $H_N$ consists of a combination of N classifiers from H,

$$\Pi_{H_N}(p) \leq (\Pi_H(p))^N \tag{4}$$

$$(\Pi_H(p))^N \leq \left(\left(\frac{ep}{d}\right)^d\right)^N \tag{5}$$

From (3), (4) and (5)

$$2^p \leq \left(\left(\frac{ep}{d}\right)^d\right)^N$$

Solving, we get the following two bounds on the value of $p$

$$p_1 = \frac{-W_0\left(\frac{-ln2}{eN}\right) \times Nd}{ln2}, p_2 = \frac{-W_{-1}\left(\frac{-ln2}{eN}\right) \times Nd}{ln2} \tag{6}$$

where $W_0(x)$ and $W_{-1}(x)$ denote the main branch and a lower branch of the Lambert W. function. Here $p_1 \leq p \leq p_2$.
The lower bound $p_1$ is a monotonically decreasing function of $N$ with a maximum value of 1.0627. The upper bound $p_2$ is a monotonically increasing function of $N$.

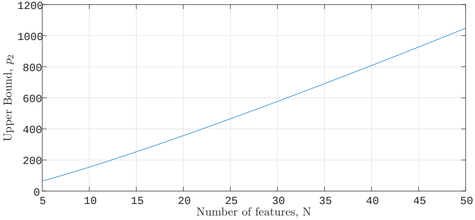



Figure : Upper bound on VC dimension $p$

# Empirical Estimation of VC dimension

- The VC dimension of any classifier space can be found by examining a plot of the ratio $(\Pi_H(m)/2^m)$ to $m$. The last value of $m$ at which the graph has the value of unity is the VC dimension of $H$.
- Growth function $\Pi_H(m)$ of the MVP classifier space was calculated by exhaustively searching through the sample space.
- The size of our search space was drastically reduced from the infinitely large real number space $\mathbb{R}^{m\times n}$ to a set consisting of $\binom{n+m!-1}{m!-1}$ inputs that are representative of all possible input combinations.
- For finding the exact value of $\Pi_{H_N}(m)$ for a given $m$ and $n$, its value was found for all $\binom{N+g-1}{g-1}$ inputs, and the maximum value among them was reported.

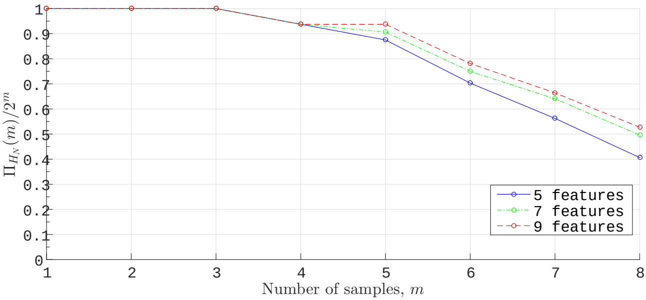



Figure : Plot of the ratio $\Pi_{H_N}(m)/2^m$ versus number of classifiers $m$. Each graph deviates from 1 at $m = 3$.

Hence it appears that the VC dimension of MVP classifier is 3, irrespective of the number of features.

# Case study on acoustic fault diagnosis

- Generalization ability of MVP classifier was compared with linear and RBF kernel SVM on acoustic data obtained from air compressors [5].
- The training and testing set both consisted of 256 samples, each with 286 features. Since the number of samples was less than the number of features, it raised the possibility of overfitting. Therefore, performance of the classifier was checked twice : 1) with all 286 features and 2) with a reduced set of 25 features.

TABLE 1:
Training & Test Accuracies

| No. of Features | MVP Classifier | | SVM with linear kernel | | SVM with RBF kernel | |
|---|---|---|---|---|---|---|
| | Train Acc. | Test Acc. | Train Acc. | Test Acc. | Train Acc. | Test Acc. |
| 286 | 100% | 99.22% | 100% | 97.66% | 100% | 76.95% |
| 25 | 100% | 99.61% | 100% | 97.66% | 100% | 89.06% |

# Conclusion

- A class of majority vote classifiers, MVP classifier was proposed that is more generalized than linear SVM on account of lower VC dimension.
- An upper bound on the VC dimension was formulated.
- The exact value was empirically estimated to be 3.
- A case study on a real world application demonstrated the high generalization ability of the MVP classifier in comparison to SVM in case of low level feature data.

# Future Work

- Checking the performance of MVP classifier on multi-class problems.
- A limitation of the MVP classifier is that in many problems it lacks sufficient flexibility to fit the training data well. Hence a possible extension of this work could involve the use of deep learning techniques for transforming low level features to higher level features, to give low training error with MVP classifier.

# References I


- Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- V. N. Vapnik, *Statistical learning theory*. Wiley New York, 1998.
- C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- N. Sauer, "On the density of families of sets," *Journal of Combinatorial Theory, Series A*, vol. 13, no. 1, pp. 145–147, 1972.
- N. Verma, R. Sevakula, S. Dixit, and A. Salour, "Intelligent condition based monitoring using acoustic signals for air compressors," *Reliability, IEEE Transactions on*, vol. PP, no. 99, pp. 1–19, 2015.