Assignment 6:
Komal Wavhal
Student Id -20034443

Summarize your understanding of VC-dimension, Natarajan dimension, and DS-dimension. Give examples to show that in binary classification, both N-dim and DS-dim reduces to the VC-dim.

Solution:

Dimension Type	**Binary Classification Equivalent**	**Generalization Type**
VC-Dimension	VC-dim	Binary classification
Natarajan Dimension	Reduces to VC-dim	Multiclass classification
Graph (DS) Dimension	Reduces to VC-dim	Multiclass (stronger)

VC-Dimension

Definition: The **VC-dimension** of a hypothesis class $\mathcal{H}$ is the largest number of points that can be **shattered** by $\mathcal{H}$, where *shattering* means that for **every** possible binary labeling of those points, there exists a hypothesis $h \in \mathcal{H}$ that realizes that labeling.

- Formally: A set $S \subseteq X$ is shattered by $\mathcal{H}$ if $\{(h(x_1), \ldots, h(x_n)) : h \in \mathcal{H}\} = \{0,1\}^n$, where $S = \{x_1, \ldots, x_n\}$.

Example: In $\mathbb{R}^2$, the class of linear separators (halfspaces) has VC-dimension 3. That means we can find 3 points in the plane that can be shattered, but no 4 points can be shattered by linear separators.

VC-Dimension (Vapnik–Chervonenkis Dimension)

The **VC-dimension** is a measure of the capacity of a hypothesis class $\mathcal{H}$ in the context of **binary classification**, where the label space is $\mathcal{Y} = \{0,1\}$ (or $\{-1,+1\}$).

- A set of instances $\{x_1, x_2, \ldots, x_d\}$ is **shattered** by $\mathcal{H}$ if, for every possible labeling of the set (i.e., all 2^d binary labelings), there exists a hypothesis $h \in \mathcal{H}$ that realizes that labeling.
- The **VC-dimension** of $\mathcal{H}$, denoted $\mathrm{VCdim}(\mathcal{H})$, is the size of the largest set that can be shattered by $\mathcal{H}$.

- The **VC-dimension** measures the largest number of points that can be **shattered** by a hypothesis class.
- A set of points is **shattered** by a hypothesis class if, for every possible labeling of the points (using 0/1 labels), there exists a hypothesis in the class that correctly classifies those labels.

Let $\mathcal{H} \subseteq \{0,1\}^X$. The VC-dimension of $\mathcal{H}$ is the size of the largest set $C \subseteq X$ such that $\mathcal{H}|_C = \{0,1\}^C$, i.e., every labeling of C can be realized.

Let $\mathcal{H}$ be the class of all thresholds on the real line:

$$\mathcal{H} = \{h_t(x) = \mathbb{I}[x \geq t] \mid t \in \mathbb{R}\}$$

- We can shatter 1 point (label it as 0 or 1 with appropriate threshold).
- We cannot shatter 2 points with all four labelings. ✅ So, VC-dim = 1.

🔑 **Theory:**

- Applies to **binary classification**.
- Measures the **capacity** of a hypothesis class $\mathcal{H} \subseteq \mathcal{X} \to \{0, 1\}$.
- A set $C = \{x_1, ..., x_n\}$ is **shattered** by $\mathcal{H}$ if for **every possible labeling** of the points (i.e., 2^n combinations), there exists a function in $\mathcal{H}$ realizing it.

📘 **Example 1: Threshold functions on the real line**

- Let $\mathcal{H} = \{h_c(x) = 1[x \geq c]\}$
- Can shatter 1 point: For $x = 0$, choose c > 0 for label 0, c < 0 for label 1.
- Cannot shatter 2 points → VC-dim = 1

📘 **Example 2: Linear classifiers in 2D**

- H: all linear decision boundaries in $\mathbb{R}^2$
- Can shatter 3 non-collinear points, not 4 in general
- So, VC-dim = 3

VC-Dimension – Examples

Example 1: Threshold Functions on the Real Line

Let $\mathcal{H} = \{h_a(x) = \mathbf{1}[x \leq a] : a \in \mathbb{R}\}$

Claim: VC-dim = **1**

- For one point x_1, we can realize both labels (0 and 1) by choosing appropriate thresholds.
- But for two points $x_1 < x_2$, we **cannot** realize all four labelings. For example, the labeling (1, 0) is not possible (if the first is 1, then the second must be too).

Example 2: Linear Classifiers in $\mathbb{R}^2$

Let $\mathcal{H}$ be all linear separators (half-planes) in $\mathbb{R}^2$.

Claim: VC-dim = **3**

- Any set of 3 non-collinear points in general position can be shattered (e.g., a triangle).
- But no 4 points can be shattered.

Natarajan Dimension

Definition: Let $\mathcal{H} \subseteq Y^X$ be a hypothesis class where Y is a finite set of labels (not just binary). A set $S = \{x_1, \ldots, x_n\} \subseteq X$ is **Natarajan-shattered** by $\mathcal{H}$ if there exist two functions $f, g : S \to Y$ such that:

1. For all i, $f(x_i) \neq g(x_i)$,
2. For **every** subset $T \subseteq S$, there exists $h_T \in \mathcal{H}$ such that:

$$h_T(x_i) = \begin{cases} f(x_i), & \text{if } x_i \in T \\ g(x_i), & \text{otherwise} \end{cases}$$

Natarajan dimension is the size of the largest set that can be Natarajan-shattered.

Natarajan Dimension

The **Natarajan dimension** generalizes VC-dimension to **multi-class classification**, where the label space $\mathcal{Y}$ has more than two classes.

- A set $\{x_1, \ldots, x_d\}$ is **Natarajan-shattered** by $\mathcal{H} \subseteq \mathcal{X} \to \mathcal{Y}$ if there exist two functions $f_1, f_2 : \{x_1, \ldots, x_d\} \to \mathcal{Y}$, such that:
  - $f_1(x_i) \neq f_2(x_i)$ for all i,
  - For every subset $S \subseteq \{x_1, \ldots, x_d\}$, there exists $h \in \mathcal{H}$ such that:

$$h(x_i) = \begin{cases} f_1(x_i) & \text{if } x_i \in S \\ f_2(x_i) & \text{if } x_i \notin S \end{cases}$$

- The **Natarajan dimension** is the size of the largest set that can be Natarajan-shattered by $\mathcal{H}$.

- Generalizes VC-dimension to **multiclass classification**.
- Measures how many points can be labeled in **two distinct ways** such that **each labeling can be realized by a different function** in the class.

Let $\mathcal{H} \subseteq Y^X$, where $|Y| \geq 2$. A set $C \subseteq X$ is N-shattered by $\mathcal{H}$ if:

- There exist **two functions** $f_1, f_2 : C \to Y$ such that $f_1(x) \neq f_2(x)$ for all $x \in C$, and
- For **every subset** $S \subseteq C$, there is an $h \in \mathcal{H}$ such that:

$$h(x) = \begin{cases} f_1(x) & \text{if } x \in S \\ f_2(x) & \text{if } x \notin S \end{cases}$$

Let $Y = \{0, 1\}$, then $f_2(x) = 1 - f_1(x)$. This reduces exactly to the VC-dimension condition: for every subset S, we can match labels (either 1 or 0), i.e., all 2^n labelings possible.

🔑 **Theory:**

- Generalizes VC-dimension to **multi-class classification** (labels in {1,...,k}).
- A set C is Natarajan-shattered by $\mathcal{H} \subseteq \mathcal{X} \to \{1, ..., k\}$ if:
  - ∃ two functions $f, g \in \mathcal{H}$, with $f(x) \neq g(x)$ for all $x \in C$
  - ∀ labeling $y : C \to \{f(x), g(x)\}$, ∃ $h \in \mathcal{H}$ such that $h(x) = y(x)$

📘 **Example 1: 3-label intervals on ℝ**

- H: assign label 1 to (-∞, a), label 2 to (a, b), label 3 to (b, ∞)
- Choose 2 points and vary the interval cutoffs → Natarajan-dim = 2

Natarajan Dimension – Examples

Example 1: All Labelings Over 2 Points in 3-Class Setting

Let $X = \{x_1, x_2\}$, $Y = \{0, 1, 2\}$, and:

$$\mathcal{H} = \{h : X \to Y\}$$

That is, the full hypothesis space. Total functions: $3^2 = 9$.

Claim: Natarajan dim = 2

Take any two different labelings f and g such that $f(x_1) \neq g(x_1)$, $f(x_2) \neq g(x_2)$. Since all possible functions exist, for any subset $T \subseteq \{x_1, x_2\}$, we can construct a function choosing labels from f on T, and from g on the rest.

Example 2: 1-vs-Rest Classifiers for 3 Classes

Suppose $X = \mathbb{R}$, $Y = \{1, 2, 3\}$, and define:

$$\mathcal{H} = \{h_a(x) = \text{argmax}_i\,(w_i x + b_i)\}$$

Where each class is assigned a linear score. This gives a standard multiclass classifier.

For specific parameters, you can Natarajan-shatter 2 points but not 3. So:

Claim: Natarajan dim = 2

DS-Dimension (Graph Dimension)

Definition: Given $\mathcal{H} \subseteq Y^X$, a set $S = \{x_1, \ldots, x_n\} \subseteq X$ is **G-shattered** (Graph-shattered) if there exists a function $f : S \to Y$ such that for every subset $T \subseteq S$, there exists $h_T \in \mathcal{H}$ such that:

- $h_T(x) = f(x)$ for $x \in T$
- $h_T(x) \neq f(x)$ for $x \in S \setminus T$

DS-dimension (a.k.a. Graph-dimension) is the size of the largest G-shattered set.

DS-Dimension (Daniely–Shalev–Shwartz Dimension / Graph Dimension)

The **DS-dimension** is another generalization of VC-dimension to multi-class classification and is **stronger** than the Natarajan dimension.

- A set $\{x_1, \ldots, x_d\}$ is **DS-shattered** by $\mathcal{H}$ if there exists a function $f : \{x_1, \ldots, x_d\} \to \mathcal{Y}$ such that:
  - For every subset $S \subseteq \{x_1, \ldots, x_d\}$, there exists $h \in \mathcal{H}$ such that:

$$h(x_i) = \begin{cases} f(x_i) & \text{if } x_i \in S \\ \neq f(x_i) & \text{if } x_i \notin S \end{cases}$$

- The **DS-dimension** is the size of the largest set that can be DS-shattered by $\mathcal{H}$.

Also a generalization of VC-dimension to multiclass settings.

Measures how many points can be labeled in one way such that **every subset** of the points can be altered (swapped to some other label) and still be realized by the hypothesis class.

The **Graph dimension (DS-dimension)** measures the richness of a hypothesis class H⊆YX in terms of how many points can be labeled such that **for every subset of those points**, the labels can be **flipped (changed)** and still be realized by some hypothesis.

Unlike the **Natarajan dimension**, which works with **two distinct labelings**, the Graph dimension fixes **one labeling** and allows label flips on subsets.

The **Graph dimension** (also called **Dudley–Sudakov dimension**) of a hypothesis class $H \subseteq \mathcal{X} \to \mathcal{Y}$ is the size of the largest set $C = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ for which:

- There exists a **base function** $f \in H$ such that:
  - For **every subset** $S \subseteq C$, there exists a hypothesis $h \in H$ satisfying:
    - $h(x) = f(x)$ for all $x \in S$
    - $h(x) \neq f(x)$ for all $x \in C \setminus S$

- You fix a labeling fff.
- For each possible **subset of inputs**, you must find an h∈Hh that:
  - **Agrees** with fff on that subset.
  - **Disagrees** with fff on the remaining points.

  So, you're effectively able to "flip" the label at any subset of the points, starting from the fixed base labeling fff.

Let's consider:

- $\mathcal{X} = \{x_1, x_2\}$
- $\mathcal{Y} = \{1, 2, 3\}$
- Let $f(x_1) = 1$, $f(x_2) = 2$

We check if for all subsets $S \subseteq \{x_1, x_2\}$, we can construct hypotheses h that agree with f on S and disagree elsewhere (using a different label from {1,2,3}).

For example:

- $S = \emptyset$: $h(x_1) \neq 1$, $h(x_2) \neq 2$
- $S = \{x_1\}$: $h(x_1) = 1$, $h(x_2) \neq 2$
- $S = \{x_2\}$: $h(x_1) \neq 1$, $h(x_2) = 2$
- $S = \{x_1, x_2\}$: $h = f$

If all 4 such functions exist in H, then DS-dim ≥ 2.

🔑 **Theory:**

- Also for multiclass classification.
- A set $C \subseteq \mathcal{X}$ is **DS-shattered** by H if:
  - ∃ function $f \in H$, such that for **every subset** $S \subseteq C$, there exists $h \in H$ where:
    - $h(x) = f(x)$ for all $x \in S$
    - $h(x) \neq f(x)$ for all $x \in C \setminus S$

This means: from a base labeling f, you can "flip" labels on any subset using some hypothesis in H.

DS-Dimension (Graph Dimension) – Examples

Example 1: Same Setup as Natarajan Example 1

Let $X = \{x_1, x_2\}$, $Y = \{0, 1, 2\}$, and:

$$\mathcal{H} = \{h : X \to Y\}$$

This is the full function class. Choose any function f, and we can construct h_T such that it matches f on any subset T, and differs elsewhere.

Claim: DS-dim = 2

Example 2: Simple Multiclass Majority Vote

Let $X = \mathbb{R}$, $Y = \{1, 2, 3\}$, and define $\mathcal{H}$ to contain classifiers that output:

- 1 if $x < 0$
- 2 if $0 \leq x < 1$
- 3 if $x \geq 1$

You can G-shatter (graph-shatter) 2 points (e.g., $-1, 0.5$) using function $f(x) = 2$, and changing labels on subsets.

Claim: DS-dim = 2

- ## Reduction to VC-Dimension (Binary Case)

- In binary case, label flips from base $f(x) \in \{0, 1\}$ can only result in 0→1 or 1→0.
- So, flipping labels on subsets of points corresponds exactly to realizing **all possible 2^n labelings**, which is the **definition of VC-dimension**.

Let X={x1,x2}, and H contains **all** binary functions from X to {0,1}:

$f(x_1)$, $f(x_2)$	Can flip on subset S	Resulting function in H?
(0,0)	S = $\{x_1\}$ → (0,0) → (1,0)? ✓	
	S = $\{x_2\}$ → (0,0) → (0,1)? ✓	
	S = ∅ → (1,1)? ✓	
	S = $\{x_1,x_2\}$ → (0,0)? ✓	

So all subset flips are possible → DS-dim = 2

But this is also the condition for **VC-dim = 2** (set $\{x_1,x_2\}$ is shattered).

✅ **Therefore, in binary classification:**

$$\text{VC-dim} = \text{Natarajan dim} = \text{DS-dim}$$

Reduction to VC-dimension in Binary Classification

Let's consider the case when $\mathcal{Y} = \{0, 1\}$, i.e., binary classification.

🔷 Natarajan Dimension reduces to VC-dimension

In the binary case, for each point x_i, the only two possible distinct labels are 0 and 1. So for any set $\{x_1, \ldots, x_d\}$, Natarajan-shattering requires that for every labeling of the points with 0 and 1, a hypothesis exists in $\mathcal{H}$ to realize it. This is exactly the definition of **shattering** in the VC-dimension sense.

➡️ **Therefore**:

$$\text{Natarajan dimension} = \text{VC-dimension} \quad \text{for binary classification}$$

🔷 DS-Dimension reduces to VC-dimension

Again, take $\mathcal{Y} = \{0, 1\}$. Let f be any fixed function (e.g., $f(x_i) = 1$ for all i). Then, DS-shattering requires that for any subset S, there exists a hypothesis that agrees with f on S and disagrees on its complement — i.e., it outputs 1 on S and 0 on the rest.

This again corresponds to realizing all possible binary labelings of the set (i.e., all subsets of the set labeled 1).

➡️ **Therefore:**

$$\text{DS-dimension} = \text{VC-dimension} \quad \text{for binary classification}$$

✅ Example to Illustrate the Reduction

Let's take a simple hypothesis class $\mathcal{H}$ over $\mathbb{R}$:

Let $\mathcal{H} = \{h_t(x) = \mathbf{1}\{x \leq t\} : t \in \mathbb{R}\}$ — the class of threshold functions.

- This class can **shatter one point**: given any single point x_1, we can assign it 0 or 1 by choosing the threshold appropriately.
- But it **cannot** shatter two points $x_1 < x_2$, because not all four binary labelings are realizable.
- So: $\text{VCdim}(\mathcal{H}) = 1$
- Since this is a **binary classification** setting:
  - $\text{Natarajan dim} = 1$
  - $\text{DS-dimension} = 1$

Relationship in Binary Classification (Y={0,1}

Let $\mathcal{H} \subseteq \{0,1\}^X$.

For VC-dimension:

We look for the largest set that can realize **all** 2^n binary labelings → i.e., every subset of the set can be represented by $\mathcal{H}$.

For Natarajan dimension:

Let $f(x) = 0$, $g(x) = 1$. Since binary labels only have 2 options, these are the only possible differing label pairs. Then the condition of Natarajan-shattering becomes:

- For every subset $T \subseteq S$, there exists $h_T \in \mathcal{H}$ such that:
  - $h_T(x) = 0$ if $x \in T$
  - $h_T(x) = 1$ if $x \notin T$

This is **equivalent** to requiring that every labeling (every subset of S) be realizable → **same as VC-dim**.

For DS-dimension:

Similar. The function $f : S \to \{0,1\}$, and for every subset T, we require:

- $h_T(x) = f(x)$ for $x \in T$,
- $h_T(x) \neq f(x)$ for $x \notin T$

This is equivalent to being able to flip values of f on **any** subset — again, all possible binary labelings → **VC-dim**.

In **binary classification**:

$$\text{VC-dim}(\mathcal{H}) = \text{Natarajan-dim}(\mathcal{H}) = \text{DS-dim}(\mathcal{H})$$

Showing That Natarajan & DS-Dimension Reduce to VC-Dim in Binary Case

Binary Example – Full Class on 2 Points

Let $X = \{x_1, x_2\}, Y = \{0, 1\}$, and $\mathcal{H} = \{h : X \to \{0, 1\}\}$, i.e., all 4 functions.

Then:

- **VC-dim** = 2 (since every subset of $\{x_1, x_2\}$ can be labeled in all 4 ways)
- **Natarajan-dim**: Use $f(x) = 0$, $g(x) = 1$. Then for every subset T, we can form a function matching f on T and g on $S \setminus T$ → exactly VC-style shattering.
- **DS-dim**: Fix any f (say, all 0s), and flip labels on arbitrary subsets → all labelings possible → again VC-style shattering.

✅ So, all three dimensions are **2**.

Dimension	General Use Case	Reduces to VC-dim in Binary Case?
VC-Dimension	Binary classification	Yes
Natarajan Dim	Multi-class classification	Yes (equivalent in binary case)
DS-Dimension	Stronger multi-class analysis	Yes (equivalent in binary case)

Binary Example – Thresholds

$$\mathcal{H} = \{h_a(x) = \mathbf{1}[x \leq a]\}$$

- VC-dim = 1
- Natarajan-dim: Only two labels; pick $f = 0$, $g = 1$. You can shatter 1 point (not 2) → same as VC-dim.
- DS-dim: Same logic, fix any label, flip on 1 point → can't do 2.

✅ Again, all dimensions = 1.

Setting	VC-dim	Natarajan-dim	DS-dim
Full binary class on 2 points	2	2	2
Thresholds in $\mathbb{R}$	1	1	1

✅ In binary classification,

$$\boxed{\text{VC-dim} = \text{Natarajan-dim} = \text{DS-dim}}$$