VC Dimension + Bayes Nets

Yucheng Low
VC Dimension
Vapnik-Chervonenkis Dimension

- A measure of the “power” or the “complexity” of the hypothesis space.

with probability at least $1-\delta$:

$$\text{error}_{true}(h) \leq \text{error}_{train}(h) + \sqrt{\frac{VC(H) \left( \ln \frac{2m}{VC(H)} + 1 \right) + \ln \frac{4}{\delta}}{m}}$$

$$\text{error}_{true}(h) \leq \text{error}_{train}(h) + \sqrt{\frac{\ln |H| + \ln \frac{1}{\delta}}{2m}}$$

It is like a substitute for $|H|$.
Higher VC dimension implies a more “expressive” hypothesis space.
Shattering

• A set of N points is shattered if there exists a hypothesis that is consistent with every classification of the N points.
Shattering

• A set of N points is shattered if there exists a hypothesis that is consistent with every classification of the N points.

Example: Linear Classifier

Shatterable?  Yes!

Yes!
Shattering

• A set of N points is shattered if there exists a hypothesis that is consistent with every assignment of the N points.

Example: Linear Classifier
VC Dimension

• Def: The maximum number of data points that can be “shattered”

If VC Dimension = d then:
1. There exists a set of d points that can be shattered
2. There does not exist a set of d+1 points that can be shattered. (or all sets of d+1 points cannot be shattered)
VC Dimension of Linear Classifier

d \geq 2
1. There exists a set of $d$ points that can be shattered
2. There does not exist a set of $d+1$ points that can be shattered
<table>
<thead>
<tr>
<th>VC Dimension of Linear Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

\[ d \geq 3 \]
VC Dimension of Linear Classifier

To show $d < 4$
There does not exist a set of 4 points that can be shattered

Quite a bit more complicated
3 Cases

Case 1: All points on a line
Not Shatterable
VC Dimension of Linear Classifier

To show $d < 4$
There does not exist a set of 4 points that can be shattered

Quite a bit more complicated
3 Cases

Case 2: Convex Hull has 3 points

Not Shatterable
VC Dimension of Linear Classifier

To show $d < 4$
There does not exist a set of 4 points that can be shattered

Quite a bit more complicated
3 Cases

Case 3: Convex Hull has 4 points
Not Shatterable
VC Dimension of Linear Classifier

- Since $d \geq 3$ and $d < 4$
- VC dimension of Linear Classifier = 3
VC dimension of Positive Triangles

**Positive Triangle Classifier:** Classifier everything inside the triangle as **positive**

With some handwaving.

\[ d \geq 7 \]
VC dimension of Positive triangles

**Positive Triangle Classifier:** Classifier everything inside the triangle as **positive**

To show $d < 8$

2 cases

Case 1: Convex hull $\leq 7$ points

Recall that inside of triangle must be positive
VC dimension of triangles

**Positive Triangle Classifier:** Classifier everything inside the triangle as **positive**

To show $d < 8$

2 cases

Case 2: Convex hull = 8 points
VC Dimension of Positive Triangle

• VC Dimension = 7

• It is quite hard to solve for VC Dimension of general triangles.
VC Dimension: Key Takeaway

If VC Dimension = d then:
1. There exists a set of $d$ points that can be shattered
2. There does not exist a set of $d+1$ points that can be shattered. (or all sets of $d+1$ points cannot be shattered)
Bayes Net
Bayes Net

• Compact representation of a probability distribution.

Vertices: Random Variables
Edges: Conditional Dependencies
“probabilistic relationships”
One CPT (Conditional Probability Table) for each variable

\[ P(\text{variable} \mid \text{parents of variable}) \]

implies the factorization:

\[
P(X) = \prod_{i=1}^{\mid X \mid} P(X_i \mid \text{parents}(X_i))
\]

\[ P(A,B,C,D) = P(A) \cdot P(B) \cdot P(C \mid A,B) \cdot P(D \mid C) \]
Independence

• **Local Markov Assumption:** A variable $X$ is independent of its non-descendents given its parents

```
  E ⊥ A | BC
  E ⊥ D | BC
  F ⊥ B | E
```

Allows you to read off some simple Independence relationships
D-Separation

• Given an independence query: Is $X \perp Y \mid Z$?
• $X$ and $Y$ are independent, if
  – All trails from $X$ to $Y$ are blocked
  – Equivalently: There is no active trail between $X$ and $Y$.

Trail: Any path from $X$ to $Y$ ignoring the direction of the arrows
Trails

Is $X \perp Y \mid Z$?

Given a trail from $X$ to $Y$: $T_1 - T_2 - T_3 \ldots - T_n$

$(T_1$ is $X$, $T_n$ is $Y$)

Blocked (inactive) if $T_i$ observed
Trails

Is $X \perp Y \mid Z$?

Given a trail from $X$ to $Y$: $T_1 - T_2 - T_3 - \ldots - T_n$

$(T_1$ is $X$, $T_n$ is $Y)$

Blocked (inactive) normally
Not blocked (active) if $T_i$ or any of its descendents are observed
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H
A trail from A to H is Active if the Bayes Ball can get from A to H
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H

V structure.
C not observed. Ball bounces away.
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H
A trail from A to H is Active if the Bayes Ball can get from A to H

V structure. C observed. Ball can pass through
A trail from A to H is Active if the Bayes Ball can get from A to H.

Ball gets stuck here
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H
Bayes Ball

A trail from A to H is Active if the Bayes Ball can get from A to H

V structure.
Descendent of F observed.
Ball can pass through
A trail from A to H is Active if the Bayes Ball can get from A to H
Bayes Ball

• Then $X \perp Y \mid Z$ if
  – All trails from $X$ to $Y$ are blocked
Test your Understanding

G ⊥ J | E ? No
H ⊥ F | G ? No
H ⊥ A | DG ? Yes
D ⊥ C | A ? Yes
D ⊥ C | A,I ? No
Teaser
Compute $P(A=T, C=T)$

\[
P(A = T, C = T) = \sum_B \sum_D P(A = T) P(B) P(C = T | A = T, B) P(D | C = T)
\]

\[
= P(A = T) \sum_B P(B) P(C = T | A = T, B) \sum_D P(D | C = T)
\]

\[
= P(A = T) \sum_B P(B) P(C = T | A = T, B) \quad \text{D disappears!}
\]
Inference

• How to do this algorithmically?
• Related to **Variable Elimination**.
• Go to next couple of lectures