Introduction to Machine Learning

CSE474/574: Lecture 5

Varun Chandola <chandola@buffalo.edu>

4 Feb 2015

Outline

Outline

1. Inductive Bias
2. Online Learning
   - Online Learning of Conjunctive Concepts
     - Properties
3. Optimal Mistake Bounds for a Concept Class
   - Bounds on the Optimal Mistake Bound

How many target concepts can there be?

- Target concept labels examples in X
- $2^{|X|}$ possibilities ($\mathcal{C}$)
- $|X| = \prod_{i=1}^{d} n_i$
- Conjunctive hypothesis space $\mathcal{H}$ has $\prod_{i=1}^{d} n_i + 1$ possibilities
- Why is this difference?

Hypothesis Assumption

Target concept is *conjunctive*.

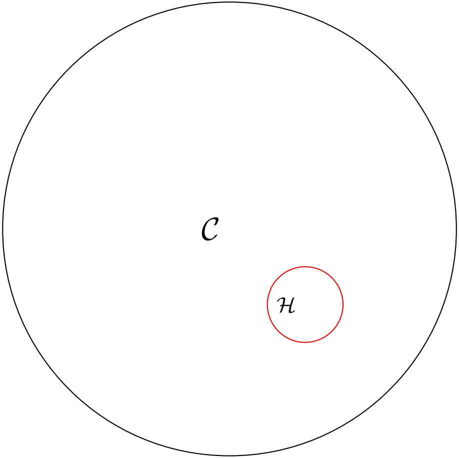
How many target concepts can there be?

- Target concept labels examples in X
- $2^{|X|}$ possibilities ($\mathcal{C}$)
- $|X| = \prod_{i=1}^{d} n_i$
- Conjunctive hypothesis space $\mathcal{H}$ has $\prod_{i=1}^{d} n_i + 1$ possibilities
- Why is this difference?

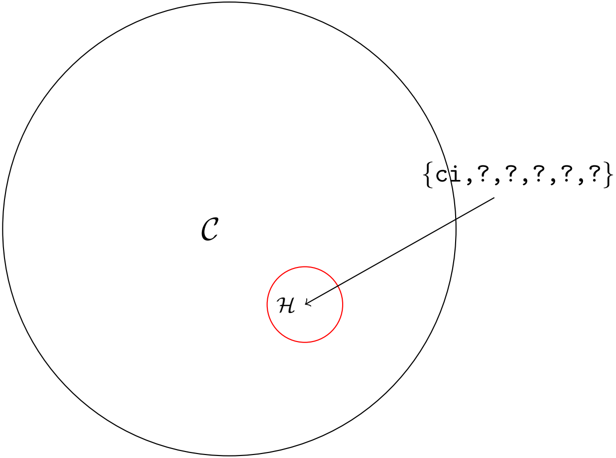
Hypothesis Assumption

Target concept is *conjunctive*.

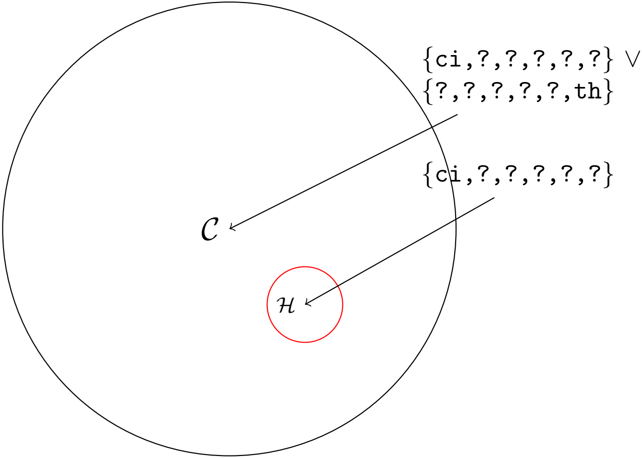
Inductive Bias



Inductive Bias



Inductive Bias



Bias Free Learning – $\mathcal{C} \equiv \mathcal{H}$

- Simple tumor example: 2 attributes - size (sm/lg) and shape (ov/ci)
- Target label - malignant (+ve) or benign (-ve)
- $|X| = 4$
- $|\mathcal{C}| = 16$

Bias Free Learning is Futile

- **A learner making no assumption about target concept cannot classify any unseen instance**

Inductive Bias

Set of assumptions made by a learner to generalize from training examples.

Examples of Inductive Bias

- *Rote Learner* – No Bias
- *Candidate Elimination* – Stronger Bias
- *Find-S* – Strongest Bias

Outline

Online Learning with Mistakes

- $X = \{true, false\}^d$
- $D = X^{(1)}, X^{(2)}, \ldots$
- $D \subseteq X$
- $c \in \mathcal{C}, c : X \rightarrow \{0, 1\}$

1: **for** $i = 1, 2, \ldots$ **do**
2: Learner given $x^{(i)} \in X$
3: Learner predicts $c_*(x^{(i)})$
4: Learner is told $c(x^{(i)})$
5: **end for**

Learning Objective

"Discover" c with minimum number of prediction mistakes

Online Learning of Conjunctive Concepts

- Target concept c is conjunctive
- Examples are denoted using binary variables v_i
  - Example: $v_1\bar{v}_2v_3v_4$
  - v_i means attribute i is *true* (or 1 or *circular*) and $\bar{v}_i$ means attribute i is *false* (or 0 or *oval*)
- Initialize $L \rightarrow \{v_1, \bar{v}_1, v_2, \bar{v}_2, \ldots, v_d, \bar{v}_d\}$
- *Match* input x and L to get prediction, $c_L(x)$
- *If* $c_L(x) \neq c(x)$ (a **mistake**)
  - Remove *offending* entries from L
- Consider next input
- L is the learnt concept when finished

Properties of Online Learning Algorithm - *Homework 1*

- Always makes mistake on the first example
- First mistake causes d entries to be removed from L
- Number of literals to be removed to reach target concept of length p is $2d - p$
- No entry in c is removed from L
- Mistakes are made only on positive examples
- Each mistake causes ≥ 1 entry to be removed from L

Mistake Bound

Concept c can be learnt with at most $d + 1$ prediction mistakes

Outline

Mistake Bound for $\mathcal{L}$

- $\mathcal{L}$ - Learning algorithm
- c - Target concept ($c \in \mathcal{C}$)
- D - One possible sequence of training examples
- $M_{\mathcal{L}}(c, D)$ - Number of mistakes made by $\mathcal{L}$ to learn c with D examples
- $M_{\mathcal{L}}(c) = \max_{\forall D} M_{\mathcal{L}}(c, D)$
- Worst case scenario for $\mathcal{L}$ in learning c
- $M_{\mathcal{L}}(\mathcal{C}) = \max_{\forall c \in \mathcal{C}} M_{\mathcal{L}}(c)$
- Worst case scenario for $\mathcal{L}$ in learning any concept in $\mathcal{C}$

Mistake Bound for $\mathcal{L}$

- $\mathcal{L}$ - Learning algorithm
- c - Target concept ($c \in \mathcal{C}$)
- D - One possible sequence of training examples
- $M_{\mathcal{L}}(c, D)$ - Number of mistakes made by $\mathcal{L}$ to learn c with D examples
- $M_{\mathcal{L}}(c) = \max_{\forall D} M_{\mathcal{L}}(c, D)$
- Worst case scenario for $\mathcal{L}$ in learning c
- $M_{\mathcal{L}}(\mathcal{C}) = \max_{\forall c \in \mathcal{C}} M_{\mathcal{L}}(c)$
- Worst case scenario for $\mathcal{L}$ in learning any concept in $\mathcal{C}$

Mistake Bound for $\mathcal{L}$

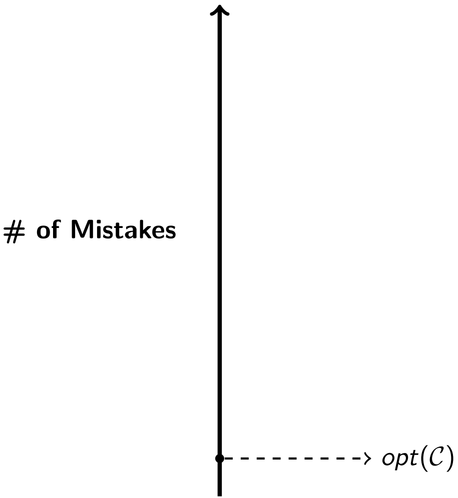
- $\mathcal{L}$ - Learning algorithm
- c - Target concept ($c \in \mathcal{C}$)
- D - One possible sequence of training examples
- $M_{\mathcal{L}}(c, D)$ - Number of mistakes made by $\mathcal{L}$ to learn c with D examples
- $M_{\mathcal{L}}(c) = \max_{\forall D} M_{\mathcal{L}}(c, D)$
- Worst case scenario for $\mathcal{L}$ in learning c
- $M_{\mathcal{L}}(\mathcal{C}) = \max_{\forall c \in \mathcal{C}} M_{\mathcal{L}}(c)$
- Worst case scenario for $\mathcal{L}$ in learning any concept in $\mathcal{C}$

Optimal Mistake Bound for Target Concept Class $\mathcal{C}$

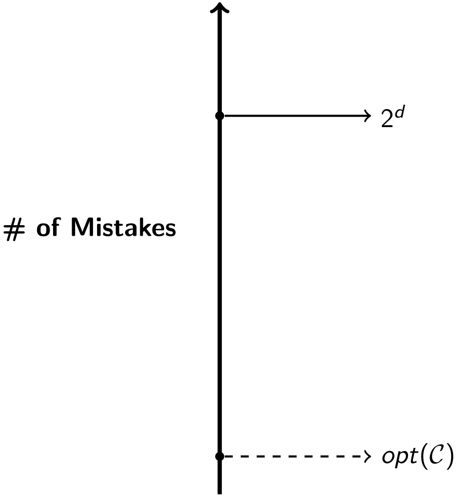
Optimal Mistake Bound

$opt(\mathcal{C}) = \min_{\forall \mathcal{L}} M_{\mathcal{L}}(\mathcal{C})$

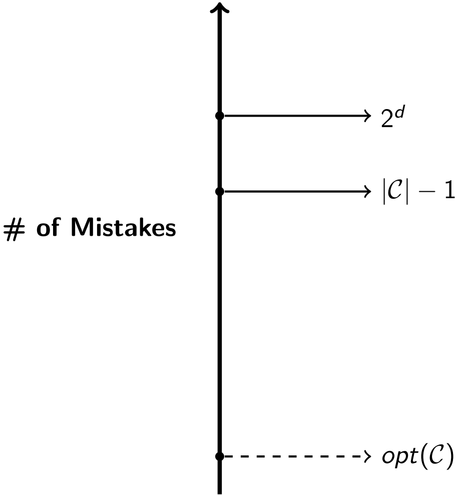
Estimating Bounds on Optimal Mistake Bound



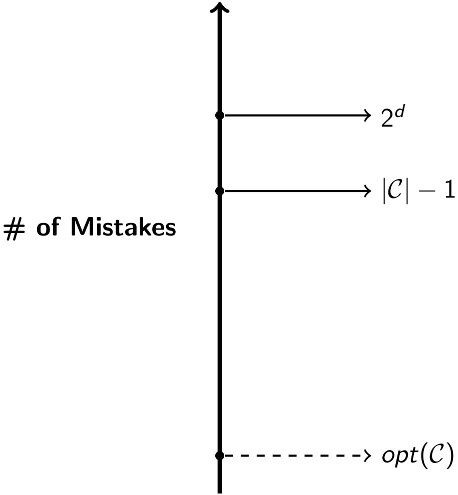
Estimating Bounds on Optimal Mistake Bound



Estimating Bounds on Optimal Mistake Bound



Estimating Bounds on Optimal Mistake Bound



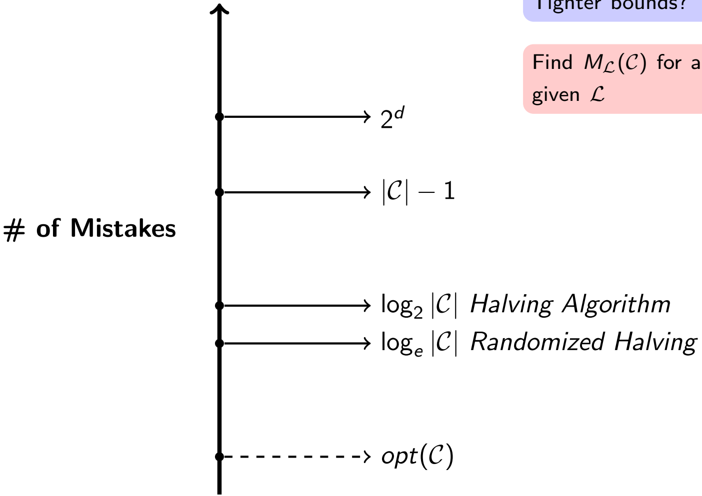
Tighter bounds?

Find $M_{\mathcal{L}}(\mathcal{C})$ for a given $\mathcal{L}$

Estimating Bounds on Optimal Mistake Bound

Tighter bounds?

Find $M_{\mathcal{L}}(\mathcal{C})$ for a given $\mathcal{L}$



References