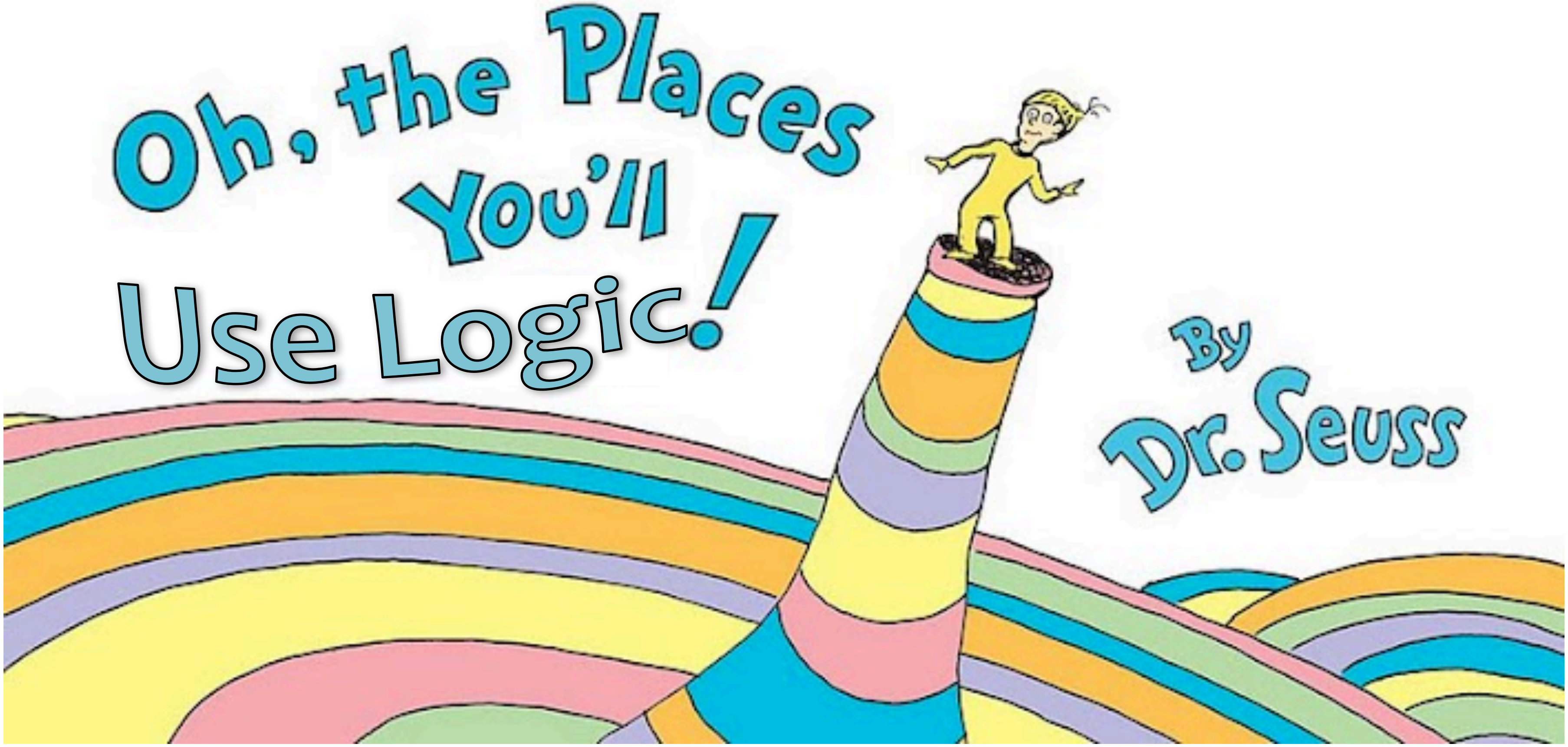
10-607 Computational Foundations for Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Propositional Logic
+
Proof Techniques

Matt Gormley
Lecture 2
Oct. 24, 2018

LOGIC


Oh, the Places You'll
Use Logic!
By Dr. Seuss

Analysis: Perceptron

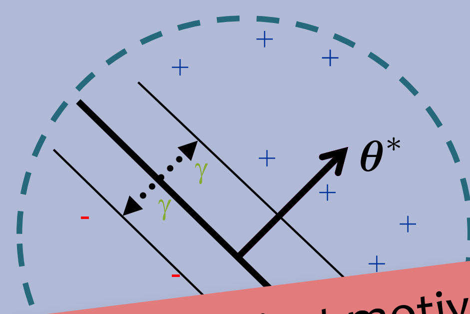
Perceptron Mistake Bound

Theorem 0.1 (Block (1962), Novikoff (1962)).
Given dataset: $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^N$.
Suppose:

1. *Finite size inputs:* $||x^{(i)}|| \leq R$
2. *Linearly separable data:* $\exists \boldsymbol{\theta}^*$ *s.t.* $||\boldsymbol{\theta}^*|| = 1$ *and* $y^{(i)}(\boldsymbol{\theta}^* \cdot \mathbf{x}^{(i)}) \geq \gamma, \forall i$

Then: The number of mistakes made by the Perceptron algorithm on this dataset is

$$k \leq (R/\gamma)^2$$



Note: This is just motivation – we'll cover the math need to understand these topics later!

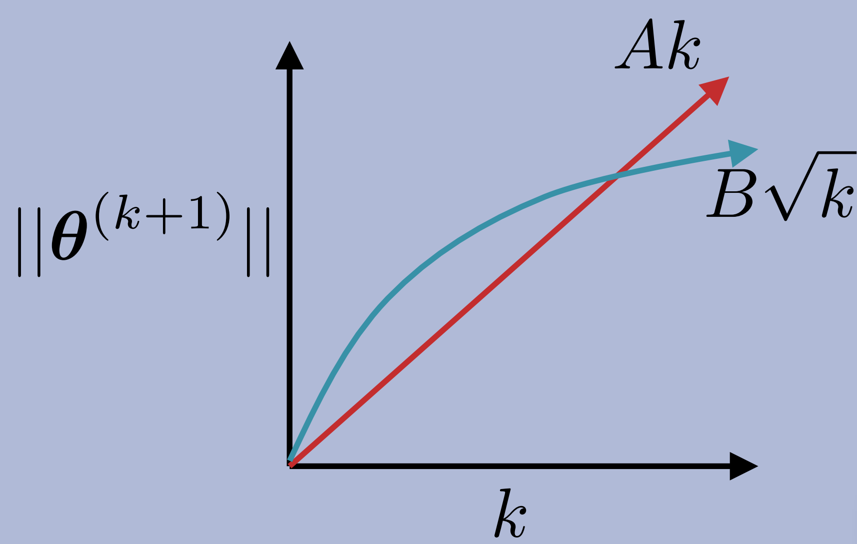
Figure from Nina Balcan

Analysis: Perceptron

Proof of Perceptron Mistake Bound:

We will show that there exist constants A and B s.t.

$$Ak \leq ||\boldsymbol{\theta}^{(k+1)}|| \leq B\sqrt{k}$$



Note: This is just motivation – we'll cover the math need to understand these topics later!

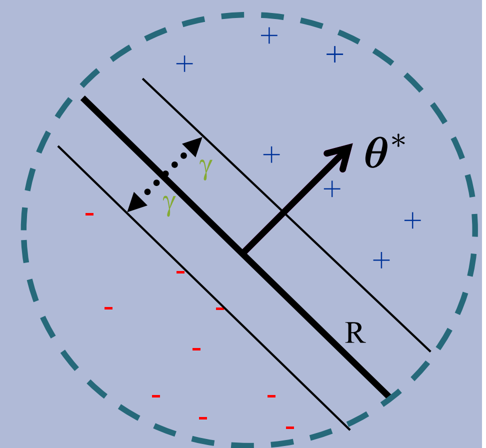
Analysis: Perceptron

Theorem 0.1 (Block (1962), Novikoff (1962)).
Given dataset: $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$.
Suppose:

1. *Finite size inputs:* $||x^{(i)}|| \leq R$
2. *Linearly separable data:* $\exists \boldsymbol{\theta}^*$ *s.t.* $||\boldsymbol{\theta}^*|| = 1$ *and* $y^{(i)}(\boldsymbol{\theta}^* \cdot \mathbf{x}^{(i)}) \geq \gamma, \forall i$

Then: The number of mistakes made by the Perceptron algorithm on this dataset is

$$k \leq (R/\gamma)^2$$



Algorithm 1 Perceptron Learning Algorithm (Online)

```
procedure PERCEPTRON(D = {(x^(1), y^(1)), (x^(2), y^(2)), ...})
    θ ← 0, k = 1                              ▷ Initialize parameters
    for i ∈ {1, 2, ...} do                    ▷ For each example
        if y^(i)(θ^(k) · x^(i)) ≤ 0 then      ▷ If mistak
            θ^(k+1) ← θ^(k) + y^(i) x^(i)     ▷ Update param
            k ← k + 1
    return θ
```

Note: This is just motivation – we'll cover the math need to understand these topics later!

Analysis: Perceptron

Proof of Perceptron Mistake Bound:

Part 1: for some A, $Ak \leq ||\boldsymbol{\theta}^{(k+1)}||$

$$\boldsymbol{\theta}^{(k+1)} \cdot \boldsymbol{\theta}^* = (\boldsymbol{\theta}^{(k)} + y^{(i)}\mathbf{x}^{(i)})\boldsymbol{\theta}^*$$

by Perceptron algorithm update

$$= \boldsymbol{\theta}^{(k)} \cdot \boldsymbol{\theta}^* + y^{(i)}(\boldsymbol{\theta}^* \cdot \mathbf{x}^{(i)})$$

$$\geq \boldsymbol{\theta}^{(k)} \cdot \boldsymbol{\theta}^* + \gamma$$

by assumption

$$\Rightarrow \boldsymbol{\theta}^{(k+1)} \cdot \boldsymbol{\theta}^* \geq k\gamma$$

by induction on k since $\theta^{(1)} = \mathbf{0}$

$$\Rightarrow ||\boldsymbol{\theta}^{(k+1)}|| \geq k\gamma$$

since $||\mathbf{w}|| \times ||\mathbf{u}|| \geq \mathbf{w} \cdot \mathbf{u}$ and $||\theta^*|| = 1$

Cauchy-Schwartz inequality

Note: This is just motivation – we'll cover the math need to understand these topics later!

Analysis: Perceptron

Proof of Perceptron Mistake Bound:

Part 2: for some B, $||\boldsymbol{\theta}^{(k+1)}|| \leq B\sqrt{k}$

$$||\boldsymbol{\theta}^{(k+1)}||^2 = ||\boldsymbol{\theta}^{(k)} + y^{(i)}\mathbf{x}^{(i)}||^2$$

by Perceptron algorithm update

$$= ||\boldsymbol{\theta}^{(k)}||^2 + (y^{(i)})^2||\mathbf{x}^{(i)}||^2 + 2y^{(i)}(\boldsymbol{\theta}^{(k)} \cdot \mathbf{x}^{(i)})$$

$$\leq ||\boldsymbol{\theta}^{(k)}||^2 + (y^{(i)})^2||\mathbf{x}^{(i)}||^2$$

since kth mistake $\Rightarrow y^{(i)}(\boldsymbol{\theta}^{(k)} \cdot \mathbf{x}^{(i)}) \leq 0$

$$= ||\boldsymbol{\theta}^{(k)}||^2 + R^2$$

since $(y^{(i)})^2||\mathbf{x}^{(i)}||^2 = ||\mathbf{x}^{(i)}||^2 = R^2$ by assumption and $(y^{(i)})^2 = 1$

$$\Rightarrow ||\boldsymbol{\theta}^{(k+1)}||^2 \leq kR^2$$

by induction on k since $(\theta^{(1)})^2 = 0$

$$\Rightarrow ||\boldsymbol{\theta}^{(k+1)}|| \leq \sqrt{k}R$$

Note: This is just motivation – we'll cover the math need to understand these topics later!

Analysis: Perceptron

Proof of Perceptron Mistake Bound:

Part 3: Combining the bounds finishes the proof.

$$k\gamma \leq ||\boldsymbol{\theta}^{(k+1)}|| \leq \sqrt{k}R$$

$$\Rightarrow k \leq (R/\gamma)^2$$

The total number of mistakes must be less than this

Note: This is just motivation – we'll cover the math need to understand these topics later!

Propositional Logic

Chalkboard

- Form of arguments
- Components of propositional logic
- Two-column proofs
- *modus ponens*
- Inference rules
- Lemmas

Inference Rules

- modus ponens: from premises ϕ and $\phi \rightarrow \psi$, conclude ψ.
- $\wedge$ introduction: if we separately prove ϕ and ψ, then that constitutes a proof of $\phi \wedge \psi$.
- $\wedge$ elimination: from $\phi \wedge \psi$ we can conclude either of ϕ and ψ separately.
- $\vee$ introduction: from ϕ we can conclude $\phi \vee \psi$ for any ψ.
- $\vee$ elimination (also called proof by cases): if we know $\phi \vee \psi$ (the cases) and we have both $\phi \rightarrow \chi$ and $\psi \rightarrow \chi$ (the case-specific proofs), then we can conclude χ.
- T introduction: we can conclude T from no assumptions.
- F elimination: from F we can conclude an arbitrary formula ϕ. (This rule is sometimes called *ex falso* or *ex falso quodlibet*, from the Latin for "from falsehood, anything.") This rule can be counterintuitive, but one way to think about it is this: we should never be able to prove F, so there's no danger in letting ourselves prove an arbitrary formula given F.
- Associativity: both $\wedge$ and $\vee$ are associative: it doesn't matter how we parenthesize an expression like $a \wedge b \wedge c \wedge d$. (So in fact we often just leave the parentheses out.)
- Distributivity: $\wedge$ and $\vee$ distribute over one another; for example, $a \wedge (b \vee c)$ is equivalent to $(a \wedge b) \vee (a \wedge c)$.
- Commutativity: both $\wedge$ and $\vee$ are commutative (symmetric in the order of their arguments), so we can re-order their arguments however we please. For example, $b \vee c \vee a$ is equivalent to $a \vee b \vee c$.

Exercise: Inference Rules

- modus ponens: from premises ϕ and $\phi \to \psi$, conclude ψ.
- $\wedge$ introduction: if we separately prove ϕ and ψ, then that constitutes a proof of $\phi \wedge \psi$.
- $\wedge$ elimination: from $\phi \wedge \psi$ we can conclude either of ϕ and ψ separately.
- $\vee$ introduction: from ϕ we can conclude $\phi \vee \psi$ for any ψ.
- $\vee$ elimination (also called proof by cases): if we know $\phi \vee \psi$ (the cases) and we have both $\phi \to \chi$ and $\psi \to \chi$ (the case-specific proofs), then we can conclude χ.
- T introduction: we can conclude T from no assumptions.
- F elimination: from F we can conclude an arbitrary formula ϕ. (This rule is sometimes called *ex falso* or *ex falso quodlibet*, from the Latin for "from falsehood, anything.") This rule can be counterintuitive, but one way to think about it is this: we should never be able to prove F, so there's no danger in letting ourselves prove an arbitrary formula given F.

- Associativity: both $\wedge$ and $\vee$ are associative: it doesn't matter how we parenthesize an expression like $a \wedge b \wedge c \wedge d$. (So in fact we often just leave the parentheses out.)
- Distributivity: $\wedge$ and $\vee$ distribute over one another; for example, $a \wedge (b \vee c)$ is equivalent to $(a \wedge b) \vee (a \wedge c)$.
- Commutativity: both $\wedge$ and $\vee$ are commutative (symmetric in the order of their arguments), so we can re-order their arguments however we please. For example, $b \vee c \vee a$ is equivalent to $a \vee b \vee c$.

Use the above inference rules to prove

$$(a \wedge b) \to (b \wedge a).$$

Write your proof in two-column format: i.e., give an explicit justification for each statement based on previous statements.

Reminder: use *only* the above rules, even if you've learned other useful rules in previous courses.

Exercise: Inference Rules

- modus ponens: from premises ϕ and $\phi \to \psi$, conclude ψ.
- $\wedge$ introduction: if we separately prove ϕ and ψ, then that constitutes a proof of $\phi \wedge \psi$.
- $\wedge$ elimination: from $\phi \wedge \psi$ we can conclude either of ϕ and ψ separately.
- $\vee$ introduction: from ϕ we can conclude $\phi \vee \psi$ for any ψ.
- $\vee$ elimination (also called proof by cases): if we know $\phi \vee \psi$ (the cases) and we have both $\phi \to \chi$ and $\psi \to \chi$ (the case-specific proofs), then we can conclude χ.
- T introduction: we can conclude T from no assumptions.
- F elimination: from F we can conclude an arbitrary formula ϕ. (This rule is sometimes called *ex falso* or *ex falso quodlibet*, from the Latin for "from falsehood, anything.") This rule can be counterintuitive, but one way to think about it is this: we should never be able to prove F, so there's no danger in letting ourselves prove an arbitrary formula given F.
- Associativity: both $\wedge$ and $\vee$ are associative: it doesn't matter how we parenthesize an expression like $a \wedge b \wedge c \wedge d$. (So in fact we often just leave the parentheses out.)
- Distributivity: $\wedge$ and $\vee$ distribute over one another; for example, $a \wedge (b \vee c)$ is equivalent to $(a \wedge b) \vee (a \wedge c)$.
- Commutativity: both $\wedge$ and $\vee$ are commutative (symmetric in the order of their arguments), so we can re-order their arguments however we please. For example, $b \vee c \vee a$ is equivalent to $a \vee b \vee c$.

Use the above inference rules to prove

$$(a \wedge b) \to (b \wedge a) .$$

Write your proof in two-column format: i.e., give an explicit justification for each statement based on previous statements.

Reminder: use *only* the above rules, even if you've learned other useful rules in previous courses.

Exercise, version 2: prove the same statement *without* using the inference rule for commutativity.

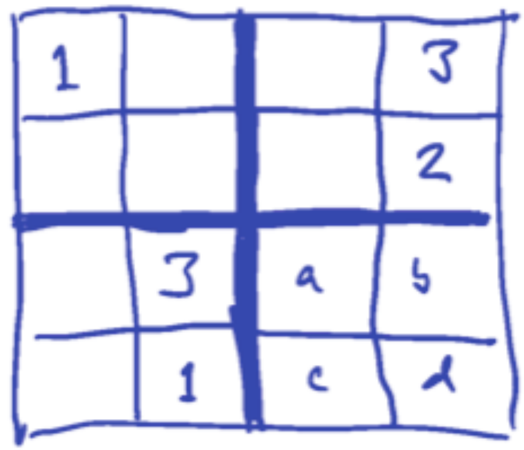
Classical Logic

Chalkboard

- Negation and constructive logic
- Law of the extended middle
- DeMorgan's laws
- Double negation elimination
- Contraposition
- Resolution
- Scoping rules

Exercise: Mini-Sudoku

In mini sudoku, the digits 1..4 must appear exactly once in each row, column, and bold-edged 2*2 box of the grid. In the grid below, we've been given five fixed digits (e.g., the 3 in the upper right corner). The squares labeled a, b, c, d are currently blank, and we'd like to figure out how to fill them in:



For example, we know that square d can't contain the digit 2, because there's already a 2 directly above it in the same column.

Fill in the squares a, b, c, d. (Note: no guessing is required.)

Use the rules of propositional logic to write down the constraints that squares a, b, c, d must satisfy. For example, you should write that the digit 1 must appear exactly once in the squares a, b, c, d. (It may take several logical formulas to implement this constraint.) For another example, you should write that the digit 2 can't appear in squares b or d (because of the 2 above them in the same column).

Prove that the solution you gave above is correct, using your formulation of the constraints together with the rules of propositional logic.

PROOF TECHNIQUES

Proof Techniques

Chalkboard

- Definitions from Discrete Math

Proof Techniques

Chalkboard

- Proof by Construction
- Proof by Cases
- Proof by Contradiction
- Proof by Contraposition
- Proof by Induction