# Optimization Techniques for Machine Learning

## AMLZC326 · #09 Duality I

Anshid Aboobacker

# WHERE WE ARE IN THE COURSE

Optimization $\Longrightarrow$ Machine Learning

Gradient Descent → Constrained Optimization → SVMs

## Unconstrained Optimization

$\min_{x} f(x)$

Examples:

- Linear regression
- Neural networks
- Gradient descent

## Constrained Optimization

$\min_{x} f(x)$ subject to $g_i(x) \leq 0$

Examples:

- SVM
- Resource allocation
- Portfolio optimization

# LEARNING OBJECTIVES

By the end of this lecture you should be able to:

- Explain why constrained optimisation is fundamentally harder than unconstrained
- Derive the Lagrangian from the indicator-function relaxation approach
- Define the primal problem as $p^* = \min_{\boldsymbol{x}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ and show it recovers the original constrained problem
- State weak duality: the dual function always provides a lower bound $d^* \leq p^*$

# TABLE OF CONTENTS

1. Why Constraints Are Hard
2. The Lagrangian
3. Primal Problem & Weak Duality

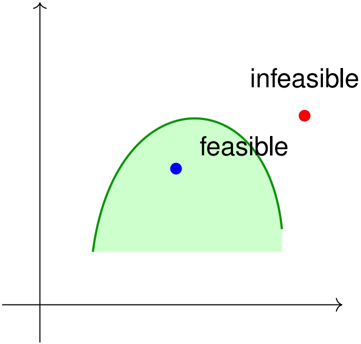
# GENERAL CONSTRAINED OPTIMIZATION PROBLEM

$$\min_x f(x) \quad \text{subject to} \quad g_i(x) \le 0 \qquad i = 1, \ldots, m$$

**Terminology**

- Feasible region
- Feasible point
- Infeasible point

$$D = \{x \in \mathbb{R}^n \mid g_i(x) \le 0\}$$



# WHY CONSTRAINTS ARE HARD

$$\min \quad f(x, y) = x^2 + y^2 \quad \text{subject to} \quad x + y \leq 6$$

**Without constraints**

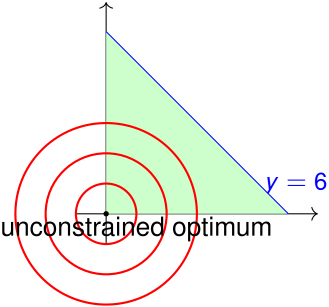
$$\nabla f = 0$$

$$(x, y) = (0, 0)$$

Easy optimization.

**With constraints**

Need to search only inside the feasible region.



# TABLE OF CONTENTS

1 Why Constraints Are Hard

2 The Lagrangian

3 Primal Problem & Weak Duality

# CONVERTING CONSTRAINTS INTO PENALTIES

$$\min_x f(x) \quad \text{subject to} \quad g_i(x) \leq 0$$

Constraint Violation $\Rightarrow$ Penalty

## Penalty Reformulation

$$J(x) = f(x) + \sum_{i=1}^{m} I(g_i(x))$$

Instead of forbidding infeasible points, we punish them heavily.

# INDICATOR FUNCTION

$$I(z) = \begin{cases} 0 & z \leq 0 \\ \\ \infty & z > 0 \end{cases}$$

**If constraint satisfied**

$$g_i(x) \leq 0$$

$$I(g_i(x)) = 0$$

No penalty.

**If constraint violated**

$$g_i(x) > 0$$

$$I(g_i(x)) = \infty$$

Infinite punishment.

Constraint violation $\Longrightarrow$ Infinite cost

# FEASIBLE VS INFEASIBLE POINTS

$$\min \quad J(x,y) = x^2 + y^2 + I(x+y-6)$$

$$x + y \leq 6$$

## Case 1: Feasible Point

Take: $(x,y) = (1,2)$
Then: $1 + 2 - 6 = -3 \leq 0$
So: $I(x+y-6) = 0$
Hence: $J(1,2) = 1^2 + 2^2 = 5$

## Case 2: Infeasible Point

Take: $(x,y) = (5,10)$
Then: $5 + 10 - 6 = 9 > 0$
So: $I(x+y-6) = \infty$
Hence: $J(5,10) = \infty$

Feasible points survive.
Infeasible points are eliminated.

# WHY INFINITE PENALTIES ARE UGLY

$$J(x) = f(x) + \sum I(g_i(x))$$

**Problem 1**

$$I(z) = \begin{cases} 0 & z \leq 0 \\ \infty & z > 0 \end{cases}$$

Not smooth.

**Problem 2**
Cannot compute gradients.

$$\nabla J(x) \quad ?$$

**Problem 3**
Gradient descent fails.

$$x_{k+1} = x_k - \eta \nabla J(x_k)$$

But: $\nabla J(x)$ is not well-behaved.

## Main Idea

We need a smooth alternative to infinite penalties.

# RELAXING THE INFINITE PENALTY

$$I(g_i(x)) \quad \Longrightarrow \quad \lambda_i g_i(x) \qquad \lambda_i \geq 0$$

Instead of assigning

$$\infty$$

to violated constraints,
we introduce a penalty weight:

$$\lambda_i$$

If the constraint violation is large, then:

$$\lambda_i g_i(x)$$

becomes large.

Smooth and differentiable.

This is the birth of the Lagrangian.

# THE LAGRANGIAN

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

**Interpretation of $f(x)$**

Original objective.

$$\min_x f(x)$$

**Interpretation of $\lambda_i g_i(x)$**

Penalty for violating constraints.

$$\lambda_i$$

measures constraint pressure.

## Key Intuition

Constraint violation increases the objective value.

# FEASIBLE CASE ANALYSIS

Suppose: $g_i(x) \leq 0$

Since: $\lambda_i \geq 0$ we get: $\lambda_i g_i(x) \leq 0$

## Observation

If: $g_i(x) < 0$
then increasing $\lambda_i$ makes $\lambda_i g_i(x)$
more negative.

## Best Choice

To maximize: $\lambda_i g_i(x)$
choose: $\lambda_i = 0$
Hence:

$$\max_{\lambda_i \geq 0} \lambda_i g_i(x) = 0$$

Feasible points receive zero penalty.

# INFEASIBLE CASE ANALYSIS

Suppose there exists a constraint such that: $g_i(x) > 0$

Since: $\lambda_i \geq 0$ we get: $\lambda_i g_i(x) > 0$

## Constraint Violated

Suppose: $g_i(x) = 10$
Then: $\lambda_i g_i(x) = 10\lambda_i$

As: $\lambda_i \to \infty$
we get: $10\lambda_i \to \infty$

## Result

$$\max_{\lambda_i \geq 0} \lambda_i g_i(x) = \infty$$

Any violated constraint produces an infinite penalty.

Constraint violation $\Longrightarrow$ Infinite punishment

# RECOVERING THE ORIGINAL PROBLEM

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

**Key Result:** $J(x) = \max_{\lambda \geq 0} L(x, \lambda)$

If $x$ is feasible:

$$J(x) = f(x)$$

No penalty.

If $x$ is infeasible:

$$J(x) = \infty$$

Constraint violation explodes.

$$\min_x J(x) = \min_x \max_{\lambda \geq 0} L(x, \lambda)$$

# TABLE OF CONTENTS

1. Why Constraints Are Hard
2. The Lagrangian
3. Primal Problem & Weak Duality

# THE PRIMAL PROBLEM

$$\boxed{\min_{x} \max_{\lambda \geq 0} L(x, \lambda)}$$

## Role of $x$

$$\min_{x}$$

Choose the best point that minimizes the objective.

$x$ tries to minimize

## Role of $\lambda$

$$\max_{\lambda \geq 0}$$

Punish constraint violations as much as possible.

$\lambda$ tries to expose violations

# A NEW QUESTION

$$\min_{x} \max_{\lambda \geq 0} L(x, \lambda)$$

What happens if we swap $\min$ and $\max$ ?

$$\boxed{\max_{\lambda \geq 0} \min_{x} L(x, \lambda)}$$

This is called the **dual problem**.

# KEY TAKEAWAYS

- Constraints make optimisation hard: feasibility and optimality must be satisfied simultaneously
- The Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_i \lambda_i g_i(\boldsymbol{x})$ converts constraints into a penalised unconstrained objective
- Primal problem: $p^* = \min_{\boldsymbol{x}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ — recovers the original constrained optimum
- Weak duality always holds: $d^* \leq p^*$ — the dual is a useful lower bound and often easier to solve

Thank you :)