Distributed optimization in Machine learning

School of Electrical and Computer Engineering

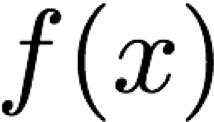
University of Tehran

Erfan Darzi

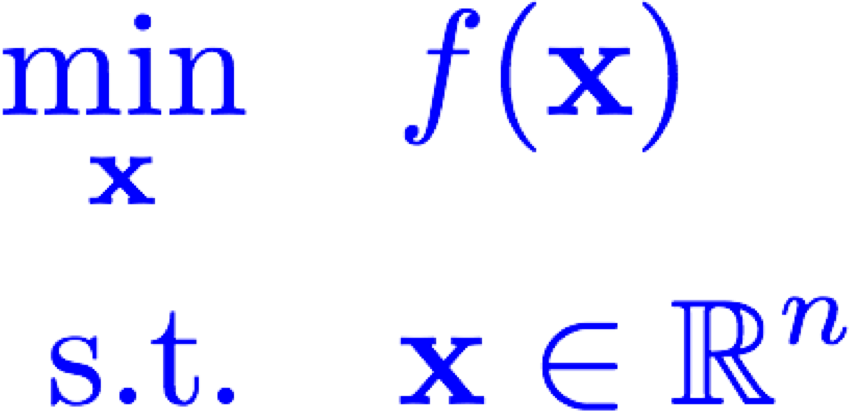
Lecture 1 – Local Optimality, Optimality Conditions, and Convex Optimization

erfandarzi@ut.ac.ir

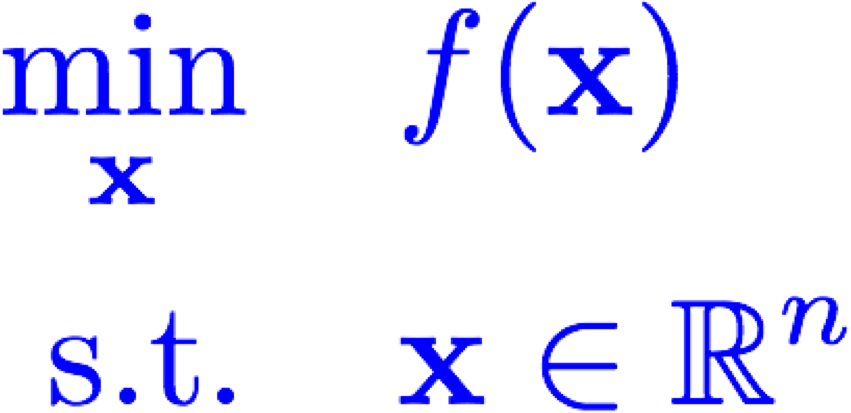
Finding a solution to an optimization problem?

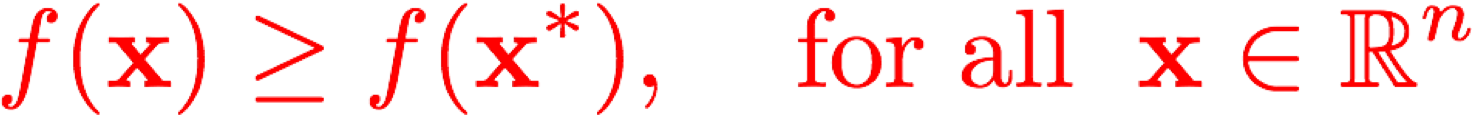
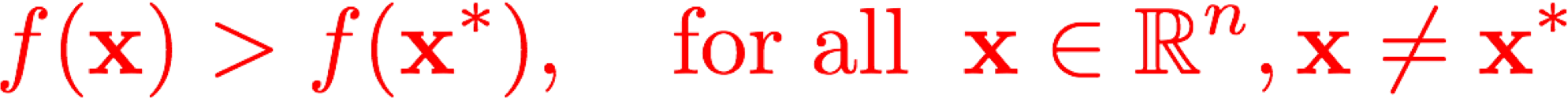


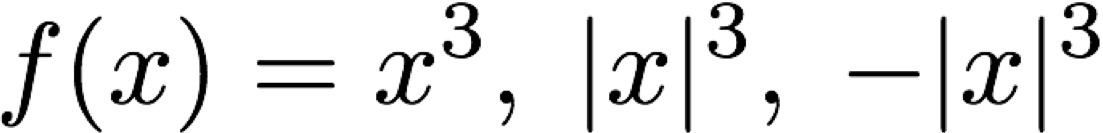
Global minimum

* Start with unconstrained case:
* Grid search?
* Requires exponential time for higher dimensions

Local minima

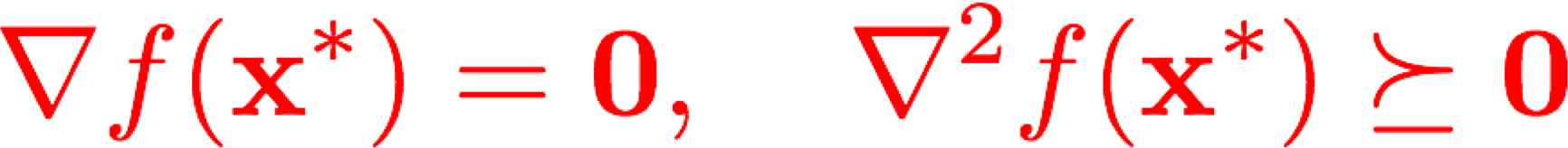
Unconstrained Optimization 

* Global minimum 
* Strict global minimum 
* Local minimum 
* Strict local minimum 

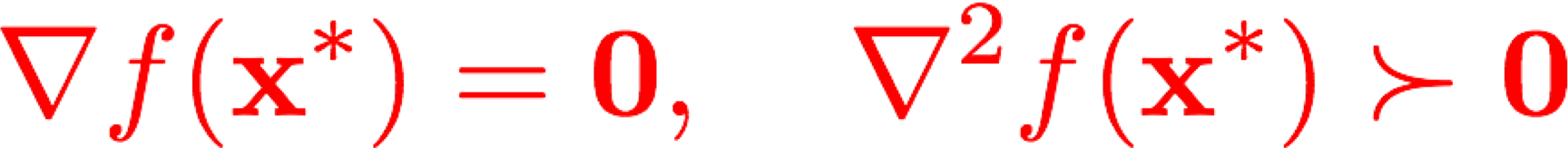
**Geometrical Interpretation! Examples:** 

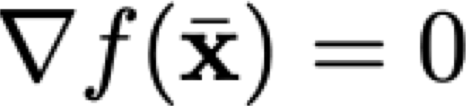
# Optimality Conditions

* Given a point , how to determine if it is a (strict) local/global minimum?
* Assume twice continuous differentiability of the objective function
* Necessary optimality condition **Is it sufficient?**



* Sufficient optimality condition for local optimality **Proofs?**



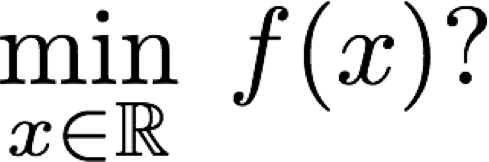
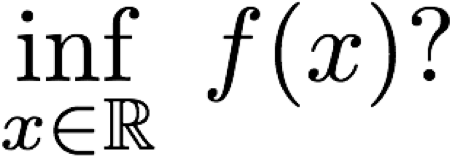
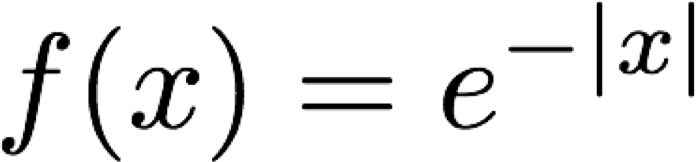
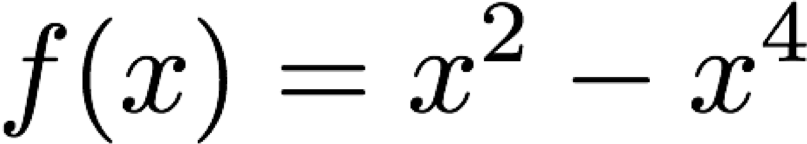
* For above unconstrained optimization, is stationary if . Figure

Why do we care?

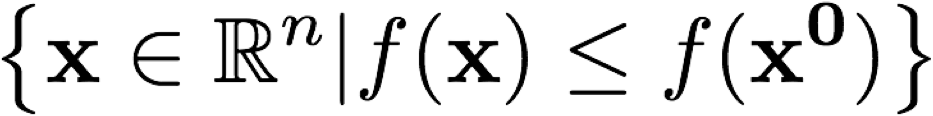
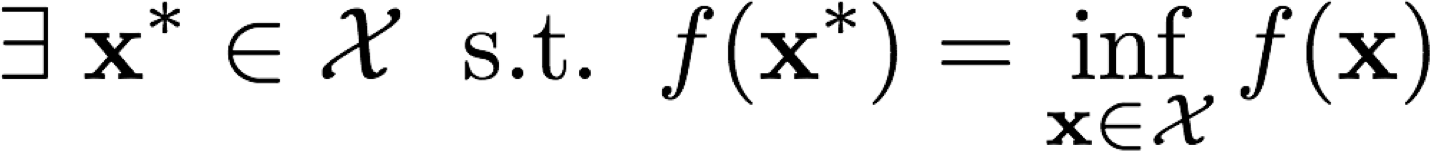
* Optimality conditions are useful because
* Tractable conditions for optimality (Yes/No)
* Narrowing down the list of potential solutions
* Useful in the **design** and the **analysis** of algorithms

Existence of An Optimal Solution

* local/global minimum/maximum of



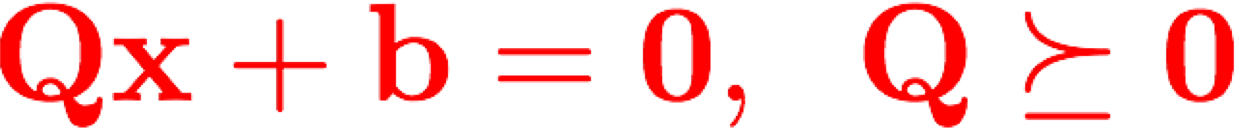
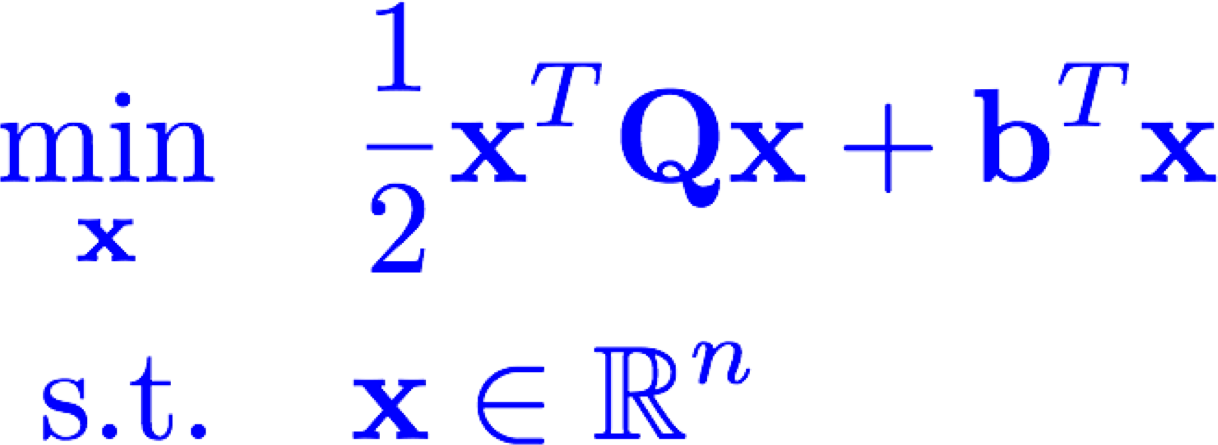
* **Bolzano-Weierstrass Theorem:** Every continuous function *f* attains its infimum over a compact set . In other words,
* Consequently, if the level set is compact, then the global min is attained



* Another sufficient condition (**coercivity**): 
* Coercivity + Continuity à existence of global optimal solution(s)

Unconstrained Quadratic Optimization

* Necessary conditions:



What if not feasible?

What if not PSD?

**Q**

is

symmeteric

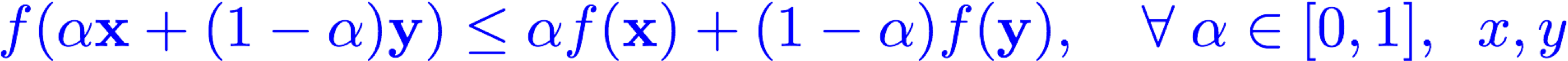
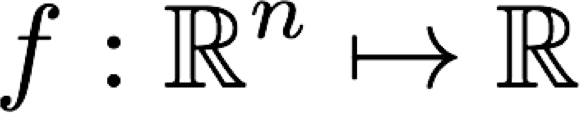
.

* Sufficient conditions?
* **Claim**:
* The above necessary condition is also sufficient
* Any local optimum is also globally optimum (True for any convex optimization)

Convexity

•

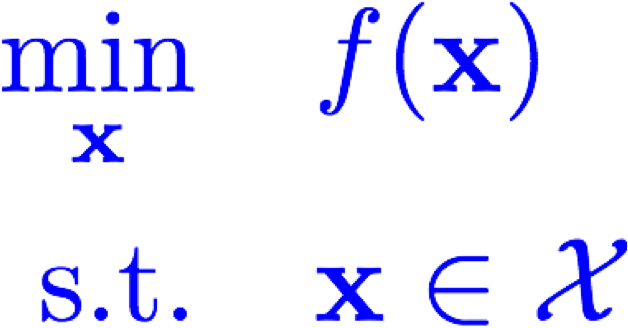
A function is convex if



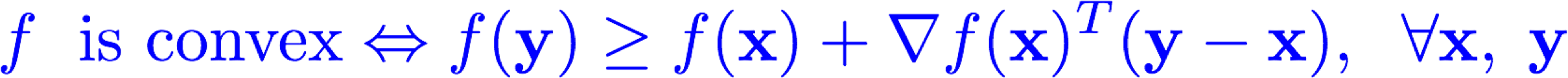
**<**

**: Strictly convex**

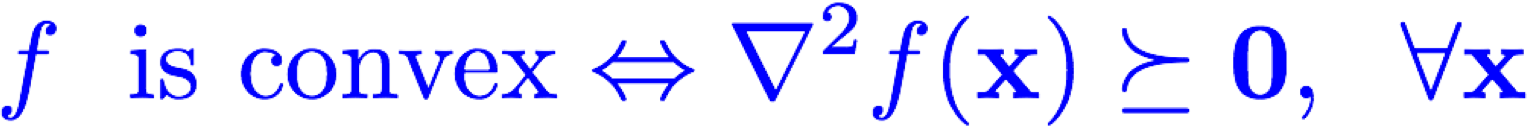
* A set is convex if its indicator function is convex
* Equivalent definition **Convex function**



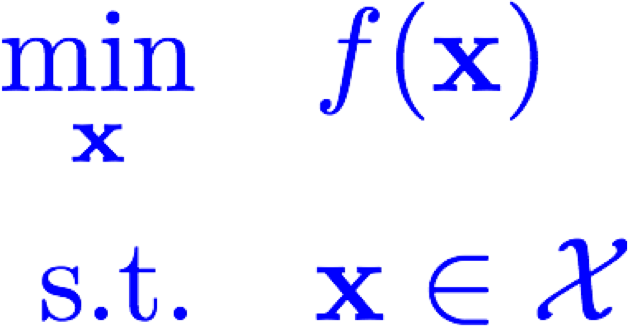
* Convex optimization problem:**Convex Set**
* **Why do we care? Local optimality (or stationarity) ==> Global optimality** Proof?
* For continuously differentiable functions Not the only class though!

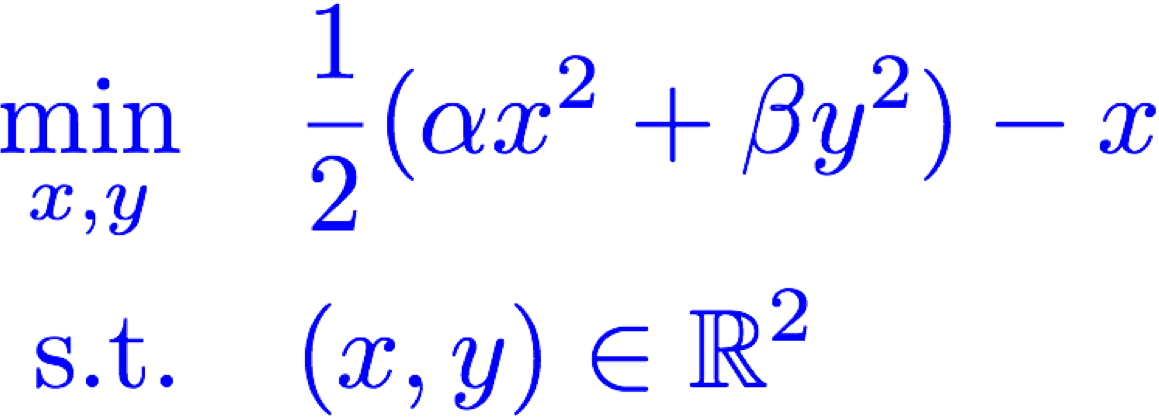


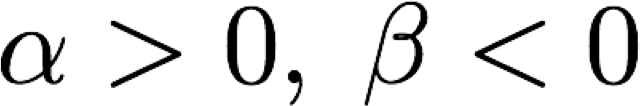
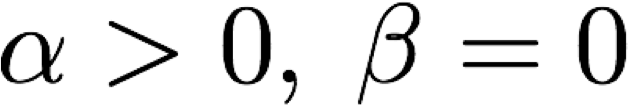
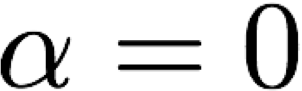
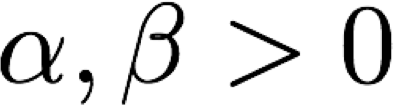
* For twice continuously differentiable functions



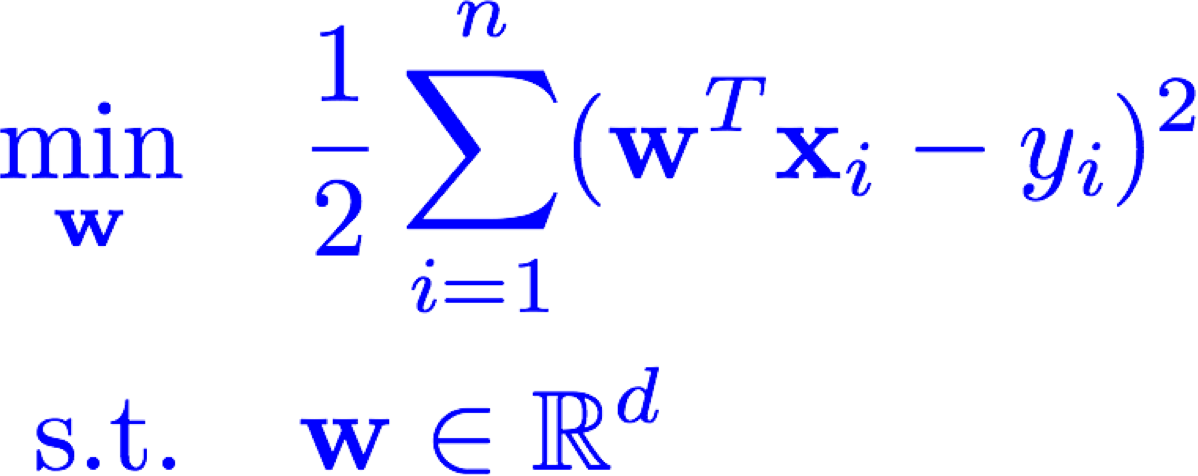
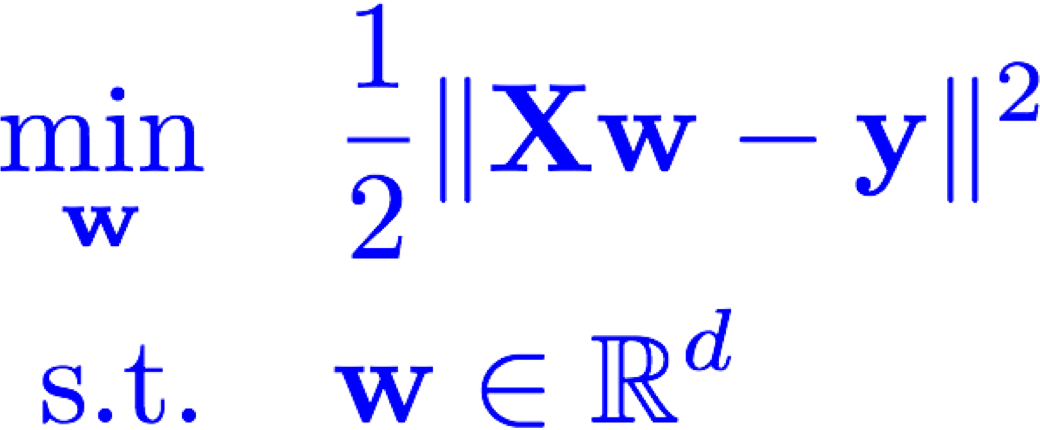
# Convexity



* **Remarks:**
* Set of optimal solutions of a convex optimization problem is convex • If the objective function is strictly convex, then the minimizer is unique
* Example: 
* Convex/Non-convex? Local/global minimum?



# Linear Regression and Linear Least Squares



* Relation to quadratic optimization
* Necessary and sufficient optimality condition: 
* **Remarks**:
* Always has a solution
* Might have unbounded level sets

# Linear Regression

**Area**

**Crime Rate**

**Age**

**RAD**

**PTRATIO**

**Bedrooms**

**…**

**Price (K)**

600

1.05

12

2.4

10.1

1

…

500

1000

2.34

10

2.5

20.1

1

…

800

1200

1.45

3

3.1

9.7

3

…

1500

1500

1.56

30

1.7

7.2

2

…

1200

…

…

…

…

…

…

…

…

2700

1.01

20

0.9

4.3

4

…

5000

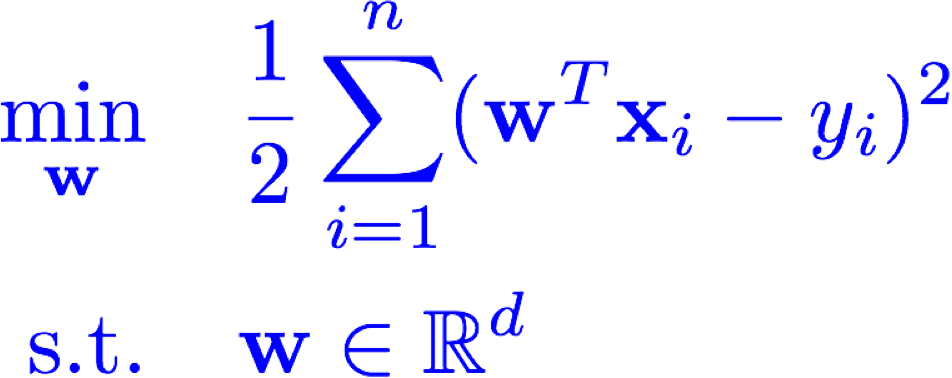


Model: Linear predictor

Loss: L2 difference

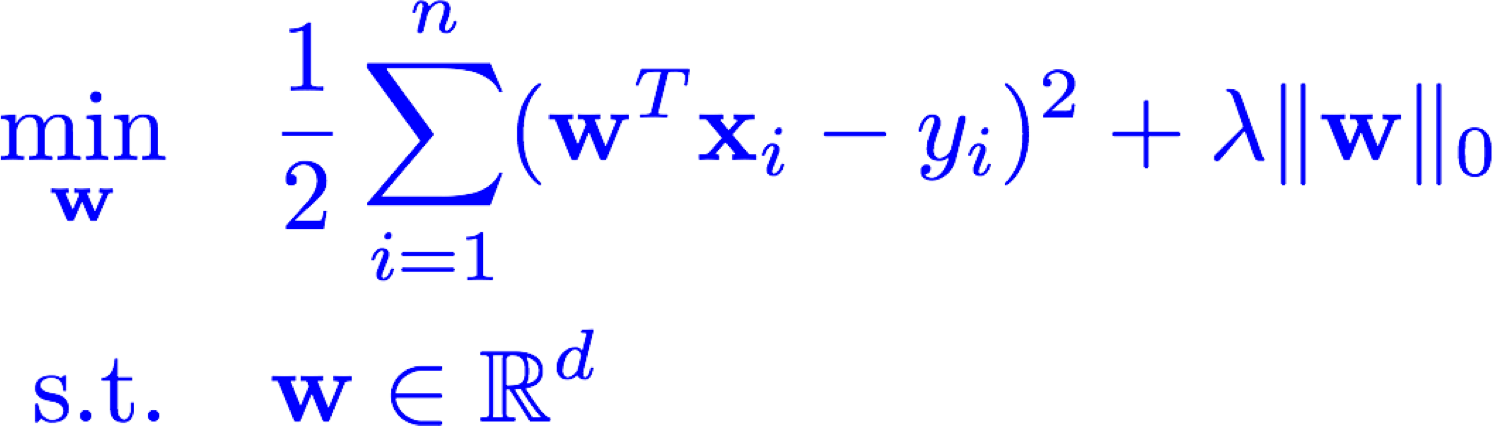
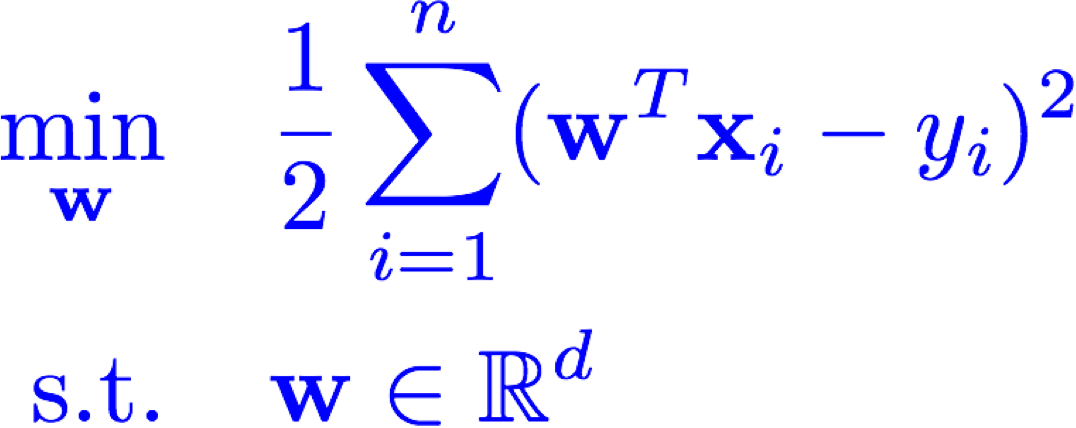


**What if we want a more interpretable predictor?**



Linear Regression

**Non-convex and even not continuous**



**Imposing sparsity**

**Suggestion**

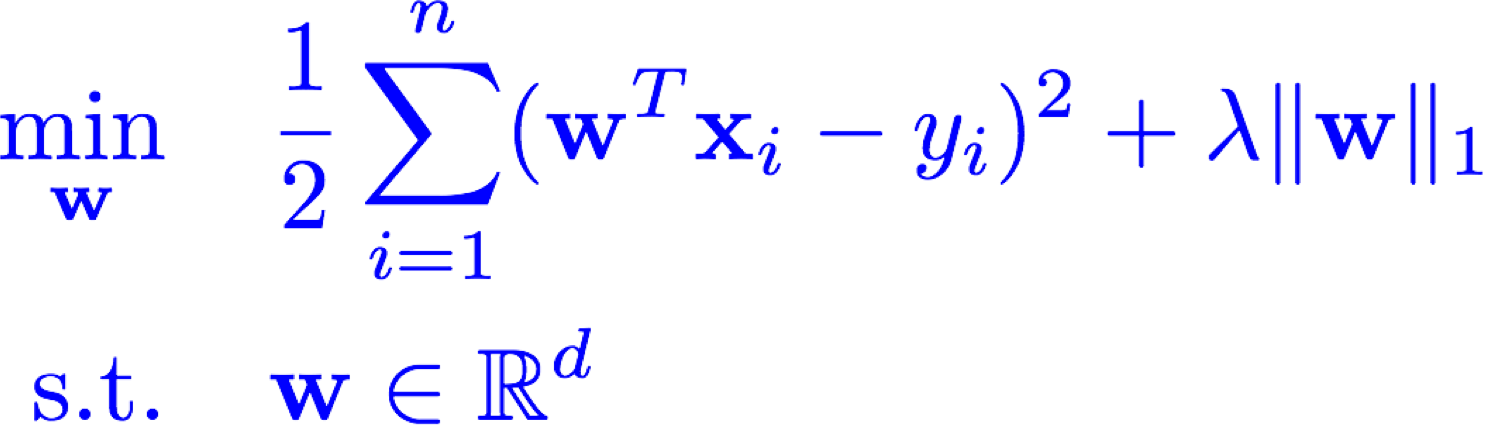
**Tibshirani**

**,**

**Donoho**

**,**

**…**



**Convex and can be solved efficiently!**

**Non**

**-**

**smooth**

**optimization problem**