Two types of optimization problems

Type I

\(n\) Items:
weights \(w_1, \ldots, w_n\),
values: \(c_1, \ldots, c_n\)

Knapsack with capacity \(W\).
Find the subset with most value that can fit into the knapsack.

Type II

\(n\) Data points (labeled):
\((x^1, y_1), \ldots, (x^D, y_D)\) where 
\(x^i \in \mathbb{R}^n\) and \(y_i \in \mathbb{R}\). Find
“best fit” linear function, i.e., find \(\beta_1, \ldots, \beta_n\) to minimize

\[
\sum_{i=1}^{D} (y_i - \beta^T x^i)^2.
\]
Two types of optimization problems

Type I

$n$ Items:
weights $w_1, \ldots, w_n$,  
values: $c_1, \ldots, c_n$  
Knapsack with capacity $W$.  
Find the subset with most value that can fit into the knapsack.

Type II

$n$ Data points (labeled):  
$(x^1, y_1), \ldots, (x^D, y_D)$ where $x^i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$.  
Find “best fit” linear function, i.e.,  
find $\beta_1, \ldots, \beta_n$ to minimize
$$
\sum_{i=1}^{D} (y_i - \beta^T x_i)^2.
$$

1. Brute force approach for Type I does not scale.
2. Classical techniques available for Type II: Calculus, convexity -  
  Does not apply to Type I
Two types of optimization problems

Type I
Combinatorial Optimization

$n$ Items:
weights $w_1, \ldots, w_n$,
values: $c_1, \ldots, c_n$
Knapsack with capacity $W$.
Find the subset with most value that can fit into the knapsack.

Type II
Continuous Optimization

$n$ Data points (labeled):
$(x^1, y_1), \ldots, (x^D, y_D)$ where $x^i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. Find “best fit” linear function, i.e., find $\beta_1, \ldots, \beta_n$ to minimize

$$\sum_{i=1}^{D} (y_i - \beta^T x^i)^2.$$  

1. Brute force approach for Type I does not scale.
2. Classical techniques available for Type II: Calculus, convexity - Does not apply to Type I
A Simpler Problem - Transshipment problem

1500

2400

3500

1200

1300

2300

1000

2000

2000
A Scheduling Problem

Job 1
Job 2
Job 3
Job 4

Machine 1
Machine 2
Machine 3
Machine 4
Machine 5
A Problem from Astronomy
A Problem from Astronomy
A Problem from Astronomy

Use physics to derive an “evaluation” function that evaluates a given partition (Correlation function in astronomy)
A Problem from Astronomy

Use physics to derive an “evaluation” function that evaluates a given partition (Correlation function in astronomy)
A Problem from Astronomy

Use physics to derive an “evaluation” function that evaluates a given partition (Correlation function in astronomy)
A Problem from Astronomy

1000 galaxies: $2^{1000}$ possible partitions
Evaluate a partition in $10^{-20}$ seconds
Will take $\sim 10^{250}$ years!!!!

Use physics to derive an “evaluation” function that evaluates a given partition (Correlation function in astronomy)
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$

Better statistical guarantees if we enforce sparsity on $\beta$. 
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$

subject to $\|\beta\|_0 \leq K$

Better statistical guarantees if we enforce sparsity on $\beta$. 
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$

subject to $\|\beta\|_0 \leq K$

Better statistical guarantees if we enforce sparsity on $\beta$.

Best Subset Selection via a Modern Optimization Lens by Bertsimas, King, Mazumder in Annals of Statistics 2016
Linear regression: Given a bunch of points $x_1, \ldots, x_D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x_i)^2$$

subject to $\| \beta \|_0 \leq K$

Better statistical guarantees if we enforce sparsity on $\beta$.

Best Subset Selection via a Modern Optimization Lens by Bertsimas, King, Mazumder in *Annals of Statistics 2016*

Similar problem in *Compressed Sensing or Sparse Coding*. 
Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$

subject to $\|\beta\|_0 \leq K$

Better statistical guarantees if we enforce sparsity on $\beta$.

Best Subset Selection via a Modern Optimization Lens by Bertsimas, King, Mazumder in *Annals of Statistics 2016*

Similar problem in *Compressed Sensing* or *Sparse Coding*.

Related problem: Robust statistics with corrupted data: Trimmed MLE, Trimmed goodness-of-fit
Statistical/Machine Learning

Linear regression: Given a bunch of points $x^1, \ldots, x^D \in \mathbb{R}^n$, and “labels” $y_1, \ldots, y_D \in \mathbb{R}$, find the best linear function that “fits” this data.

$$\min_{\beta \in \mathbb{R}^n} \sum_{i=1}^{D} (y_i - \beta^T x^i)^2$$

subject to $\|\beta\|_0 \leq K$

Better statistical guarantees if we enforce sparsity on $\beta$.

Best Subset Selection via a Modern Optimization Lens by Bertsimas, King, Mazumder in Annals of Statistics 2016

Similar problem in Compressed Sensing or Sparse Coding.

Related problem: Robust statistics with corrupted data: Trimmed MLE, Trimmed goodness-of-fit

Check out: Dimitris Bertsimas and Rahul Mazumder at MIT.