Lecture #23

Last Time: Compressed Sensing

(Recovering almost k-sparse vectors in $O(k \log \frac{n}{k})$ measurements)

Today: Smoothed Analysis

worst-case analysis often too pessimistic, and average-case analysis sensitive to distribution

Smoothed Analysis: [Spielman, Teng]

$$\max_{x} \mathbb{E}_{\sigma}\left[\text{time}(Alg(x+\sigma))\right]$$

Gaussian perturbation

Three approaches to LPs: (1) Simplex (2) Ellipsoid (3) Interior Point

exponential worst-case, but usually much better

Theorem [Spielman, Teng] The Simplex method runs in smoothed polynomial time

Smoothed analysis has been applied to: (1) mathematical programming (2) numerical analysis (3) learning (4) approximation algorithms, etc

---

Today we will cover knapsack, following [Beier, Vöcking]

$$\max \sum x_i \underset{\text{value}}{v_i} \quad \text{s.t.} \quad \sum x_i \underset{\text{weights}}{w_i} \leq W$$

$$x_i \in \{0,1\} \quad \forall i$$

Knapsack is NP-hard, but often easy

Nemhauser-Ullman Algorithm

Set $P_0 = \phi$

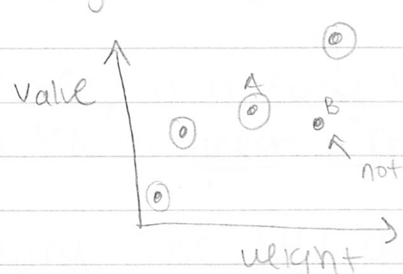
For $i = 1$ to n

  Let T $\{A\}_{A \in P_{i-1}} \cup \{A \cup \{i\}\}_{A \in P_{i-1}}$

  Remove sets from T if strictly dominated by any other

Find $A \in P_n$ with $\sum_{i \in A} w_i \leq W$, that maximizes $\sum_{i \in A} v_i$

This algorithm constructs Pareto curves, e.g.



not on Pareto curve because

$\sum_{i \in B} w_i > \sum_{i \in A} w_i$

$\sum_{i \in B} v_i \leq \sum_{i \in A} v_i$

A strictly dominates B; ($>$ and $\leq$) OR ($\geq$ and $<$)

Lemma: Each P_i is the Pareto curve for $2^{[i]}$

Proof: By induction P_{i-1} is the Pareto curve for $2^{[i-1]}$

Consider $B \subseteq [i]$ with $i \in B$. Then if $B \setminus \{i\} \notin P_{i-1}$, B cannot be on P_i because

If $A \subseteq [i-1]$ and A strictly dominates $B \setminus \{i\}$ then $A \cup \{i\}$ strictly dominates B.

Thus all feasible candidates for P_i are considered,

$\Rightarrow P_i$ is Pareto curve for $2^{[i]}$ ∎

Corollary: NU-Algorithm returns optimal soln

Proof: P_n is the Pareto curve for $2^{[n]}$ ∎

When is the NU-algorithm efficient?

Worst-case: $|P_n| \geq c^n$

Theorem: (informal) [Beier-Vöcking] The expected size of each Pareto curve P_i is polynomial, in the smoothed analysis model.

Moreover, it is easy to see that P_i can be computed in linear time from P_{i-1}

Corollary: The NU-algorithm runs in expected smoothed polynomial time

---

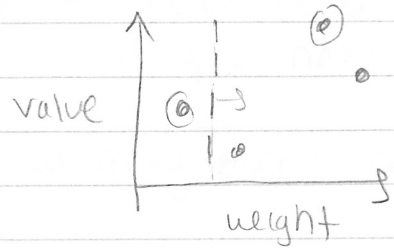
Now let's define the relevant smoothed model:

- Let z_i be independent r.v.s whose pdf is bounded by ϕ, supported in $[0,1]$
- let $v_i = v_i' + z_i$, $v_i' \in [0,1]$ worst case
- Let w_i be worst-case, arbitrary but distinct

Our goal is to build up a family of events that will let us bound the size of the Pareto Curve (PO)

Step # 1: A definition of Pareto optimal, via sweeping

Consider sweeping from low to high weight:



Observation: A point X is Pareto optimal iff when it arrives, it has strictly largest value

Aside: If 2^n points had Gaussian values (average case rather than smoothed) we immediately have

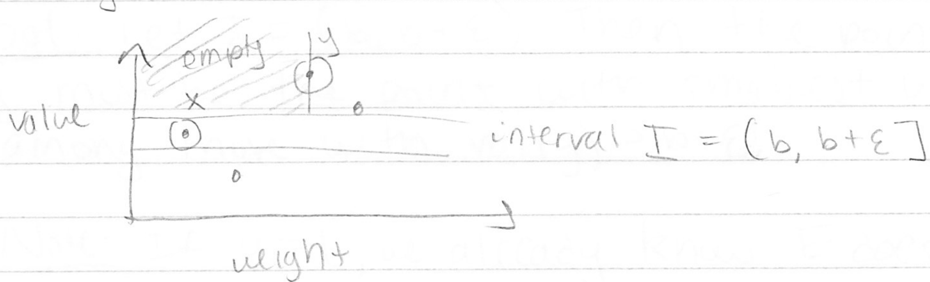
unstructured

$$\#[|PO|] = \sum_{i=1}^{2^n} \frac{1}{i} \approx n \ln 2$$

Step #2: Find an event to blame, when $x \in PO$

- Divide $[0, 2n]$ into intervals of width ε,

Now if $x \in PO$, can continue sweep and find next point $y \in PO$



Let i be a coordinate s.t $X_i \neq Y_i$. To keep things simple, suppose $x_i = 1$, $y_i = 0$ (other case is basically the same).

Now we are ready to define the family of events E, specified by interval I and index i, and a bit $\underline{a}$

$E \triangleq$ There is an $x \in PO$ with $x \in I$, and if y is next point on PO, $x_i = a$, $y_i = \bar{a}$

How many events are there? $\left(\frac{2n}{\varepsilon}\right)(n)(2) = \frac{4n^2}{\varepsilon}$

<u>Claim:</u> If no two points land in same interval,

$$|PO| \leq \sum_E \mathbb{1}_E + 1$$

last point on PO, with no y

<u>Step #3:</u> Bound the probability of each event.

Let's consider $x_i = 1$, $y_i = 0$ case, and do some <u>backwards</u> reasoning:

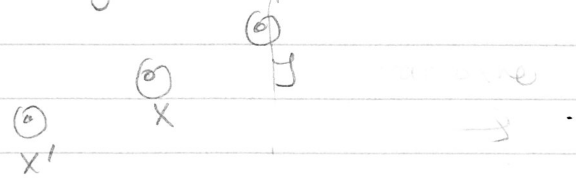
<u>Lemma:</u> If $v_1, v_2 \ldots v_{i-1}, v_{i+1} \ldots v_n$ are fixed, there is a unique x that can cause E

<u>Proof:</u> Let $I = (b, b+\varepsilon)$. Then the point y must be the point with smallest weight, among those with $val(y) > b + \varepsilon$.

<u>Note:</u> If $y_i = 1$, we already know E does not occur.

Now x is the point among those with $x_i = 0$, $weight(x) < weight(y)$ that has largest value.

Why? For any other point x' we have



and if $x' \in I$, then $val(x) > b + \varepsilon$ (no two points in same interval), but all other points y' with $val(y') > b + \varepsilon$ and $y'_i = 1$ are right of y.

Thus the next PO point after x' cannot have i^{th} coordinate equal to zero, so E does not happen. $\blacksquare$

To finish, V_i is still random so at most $\phi\varepsilon$ chance x lands in i. Thus:

<u>Lemma:</u> $\Pr[E] \leq \varepsilon\phi$

Putting it all together we have: ($\varepsilon \to 0$, so no two in same interval a.s.)

$$\mathbb{E}[|PO|] \leq 4n^2\phi + 1$$

The exciting takeaway is that the <u>explanatory power</u> of Theory is not necc. limited to worst case or average case.

When faced with a hard problem, explore it in weaker models