15-750
2/6/19

Dynamic Programming

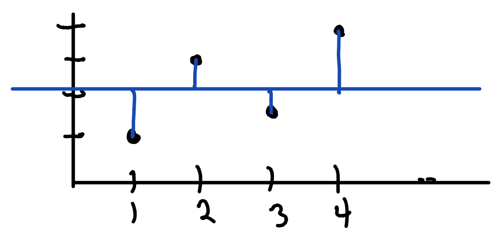
First Example: Finding a function that best fits data.

Data: $P_1=(1,Y_1), \dots, P_n=(n,Y_n)$ for $Y_i \in \mathbb{R}$.
this is 1D data

More general data may be 2D Images
3D Medical Images
d-D bio ML

Lets fit data with constant fcn.
We need at cost, penalty, or fidelity.
Lets start with quadradic cost. $\min_{a\in\mathbb{R}} \sum_{i=1}^{n}(a-Y_i)^2 = C(a)$

eg $Y=(1,3,2,4)$



Claim $a = \text{aver}(Y) = (\sum Y_i)/n$

pf Take derivative & set to zero.

$C' = \sum 2(a-Y_i) = 0$

$n\cdot a = \sum Y_i \Rightarrow a = \sum Y_i / n$

Suppose we want to have a online running average: $a_i = \sum_{j=1}^{i} Y_j / i$

We use moments

<u>Def</u> kth moment $M_k = \sum_{i=1}^{n} Y_i^k \quad k \in \{0, 1, 2, \ldots\}$

eg $M_0 = \sum_{i=1}^{n} Y_i^0 = n$ thus $\text{aver} = \dfrac{M_1}{M_0}$

Our cost fcn is just the standard deviation.

$$\sigma^2(Y_1, \ldots, Y_n) \overset{\text{def}}{=} \sum_{i=1}^{n} \left(\frac{M_1}{M_0} - Y_i\right)^2$$

$$= \sum_{i=1}^{n} \left(\frac{M_1}{M_0}\right)^2 - 2\frac{M_1}{M_0} \sum_{i=1}^{n} Y_i + \sum Y_i^2$$

$$= M_0 \left(\frac{M_1}{M_0}\right)^2 + 2\,\frac{M_1^2}{M_0} + M_2 = M_2 - \frac{M_1^2}{M_0}$$

Thus if we maintain M_2, M_1, M_0 we can compute σ^2 in constant more work.

<u>Fact</u>: $\sigma^2 = M_2 - M_1^2 / M_0$

We will use this fact later!

Consider fitting with piecewise constant but with steps or breakpoints.

Suppose we charge $\alpha \in \mathbb{R}^+$ for each step.

<u>Def:</u> $\delta(a,b) = \begin{cases} 0 & \text{if } a=b \\ 1 & \text{o.w.} \end{cases}$

<u>Goal:</u> Find $f = \langle f_1, \dots, f_n \rangle$ $f_i \in \mathbb{R}$

to $$\min_f \sum_{i=1}^{n} (f_i - Y_i)^2 + \alpha \sum_{i=1}^{n-1} \delta(f_i, f_{i-1})$$

Such problems can be solved using <u>Dynamic Programming</u>

We will think of DP as a 4-step process.

<u>Step 1</u> (def set of subproblems)

$C(j) \stackrel{\text{def}}{=} \min\limits_{f_1 \dots f_j} \text{Penalty}(f_1, \dots, f_n) \equiv$ cost of opt for $(Y_1, \dots, Y_j)$

<u>Step 2</u> (def recurrence)

Base Case: $\bar{C}(1) = 0$

$$\bar{C}(j) = \min \begin{cases} \sigma^2(Y_1, \dots, Y_j) & (*) \\ \min\limits_{1 \le i < j} \{ \bar{C}(i) + \alpha + \sigma^2(Y_{i+1}, \dots, Y_j) & (**) \end{cases}$$

Intuition:

(*) Try fitting $y_1, \dots, y_j$ with constant fcn

(**) Try last step at i

Step 3 (Show reecurence is correct) ($C = \bar{C}$)

Claim $C(j) = \bar{C}(j)$

Since Step2 finds an f and correctly

computes its we get $\forall j \; C(j) \le \bar{C}(j)$

to show $\bar{C}(j) \le C(j)$

proof Use induction on j.

Base case $j=1$ is OK.

Assume $j>1$ & $\forall i \; 1 \le i < j \quad \bar{C}(i) \le C(i)$

Suppose opt solution is $f_1, \dots, f_j$

Case1 f has no steps ie $f_1 = f_j$

$$\text{Penalty}(f_1, \dots, f_j) \geq \sigma^2(y_1, \dots, y_j) \geq \bar{C}(j)$$

Case2 Last step of f is at i

ie $f_i \ne f_{i+1}$ & $f_{i+1} = f_{i+2} = \cdots f_j$

$$\begin{aligned} \text{Penalty}(f) &\geq \text{Penalty}(f_1, \dots, f_i) + \alpha + \sigma^2(y_{i+1}, \dots y_j) \\ &\geq \bar{C}(i) + \alpha + \sigma^2(y_{i+1}, \dots, y_j) \\ &\geq \bar{C}(j) \end{aligned}$$

Step 4 (determine runtime)

1) # subproblems is n
2) Cost per subproblem $= O(n^2)$

Total $O(n^3)$

Note: Main cost was computing variances.
$\sigma^2(Y_i, \ldots, Y_j) \; \forall i < j$.

Goal: Compute the n^2 variances in $O(n^2)$ time

Idea: Make table of moments M_1, M_2

eg DP $\equiv$ $M_k(Y_i, \ldots Y_j) = \begin{cases} Y_j^k & \text{if } i = j \\ Y_j^k + M_k(Y_i, \ldots Y_{j-1}) & \text{O.W.} \end{cases}$

Use Formula $\sigma^2(\,) = M_2 - M_1^2 / M_0$

Total Cost $O(n^2)$

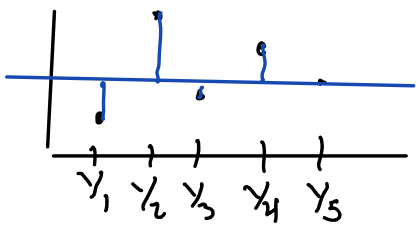
Open Question: Find $o(n^2)$ alg for piecewise constant.

The L_1 case

Prob: Input $Y = \langle y_1, \dots, y_n \rangle$

Goal: $$\min_f \sum_{i=1}^{n} |f_i - y_i| + \alpha \sum_{i=1}^{n-1} \delta(f_i, f_{i+1})$$

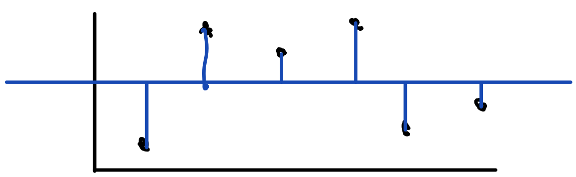
Consider no step solution:



If n is odd

$f \equiv \text{median}\{y_1, \dots, y_n\}$

Suppose n is even



All f between 2 medians work. Same fidelity

Def: $\text{Median}\{y_1, \dots, y_n\} = \{a \mid |\{y_i < a\}| = |\{y_i > a\}|\}$

(Penalty)

$$\rho(y_i, \dots, y_j) = \sum_{t=i}^{j} |y_t - a| \quad \text{where } a \in \text{median}\{y_i, \dots, y_j\}$$

Step 1 $C(j) = \min \text{cost for } \{y_1, \dots, y_j\}$

<u>Step2</u> $\bar{C}(1) = 0$

$$\bar{C}(j) = \min \begin{cases} \rho(y_1, \ldots, y_j) \\ \min\limits_{i<j} \{ \bar{C}(i) + \alpha + \rho(y_{i+1}, \ldots, y_j) \end{cases}$$

<u>Step3</u> Correctness $(C(j) = \bar{C}(j))$

Same as last time

<u>Step4</u> Timing

$\#$subprobs $= O(n)$

Given take of values $\rho(i,j) = \rho(y_i, \ldots, y_j)$.

As last time

each $\bar{C}(j)$ is $O(j) = O(n)$ time.

Total time is $O(n^2)$ plus cost of $P(i,j)$ table.

<u>Prob</u>: Input $\langle y_1, \ldots, y_n \rangle$

Compute table $\rho(i,j)$ $\forall i \le j$.

<u>Step1</u> $\rho(i,j)$ $\forall i \le j$

Step 2 Recurrence

$\tilde{\rho}(i,j) = 0$ if $i=j$

Set $Y_m = \begin{cases} \text{median}\{Y_i, \cdots Y_j\} & \text{if } j-i+1 \text{ is odd} \\ \text{median}\{Y_i, \cdots Y_{j-1}\} & \text{O.W} \end{cases}$

Set $\bar{\rho}(i,j) = \rho(i, j-1) + |Y_j + Y_m|$

Correctness Base case OK

Assume true for $\rho(i,j) = \bar{\rho}(i,j)$ for $j-i < K$

Let $j-i = K$

Case 1 $j-i+1$ odd then $Y_m = \text{med}(i,j)$

note $Y_m \in \text{med}(i, j-1)$ thus

$$\rho(i,j) = \sum_{t=i}^{j} |Y_m - Y_t| = \sum_{t=i}^{j-1} |Y_m - Y_t| + |Y_j - Y_m|$$

$$= \rho(i,j-1) + |Y_j - Y_m| \geq \bar{\rho}(i,j-1) + |Y_i - Y_m|$$

$$\geq \bar{\rho}(i,j)$$

<u>Case 2</u> $j-i+1$ even then $y_m = \mathrm{med}(i,j-1)$

note $y_m \in \mathrm{med}(i,j)$

$$\rho(i,j) = \sum_{t=i}^{j} |y_m - y_t| = \sum_{t=i}^{j-1} |y_m - y_t| + |y_j - y_m|$$

$$= \rho(i,j-1) + |y_j - y_m| \geq \bar{\rho}(i,j-1) + |y_j - y_m|$$

$$\geq \bar{\rho}(i,j)$$

<u>HW</u> Table of medians can be computed in $O(n^2 \log n)$

Important other objective functions

Total Variation $L_2 - L_1$

$$\min_f \left\{ \sum (f_i - y_i)^2 + \alpha \sum |f_i - f_{i+1}| \right\}$$

4 possible types

$L_2^2 - L_2^2$, $L_2^2 - L_1$, $L_1 - L_2$ $L_1 - L_1$

most used

TV used in image denoising

Kleinberg-Tardos does case of Piecewise linear.

Total Variation over graphs

$G = (V, E)$ $f : V \longrightarrow \mathbb{R}$

$$\min_f \left\{ \sum_{v \in V} (f(v) - g(v))^2 + \sum_{(u,v) \in E} |f(u) - f(v)| \right\}$$