Submodular functions:

- let N be a finite set
- 2^N represents the set of all subsets of N

a function $f: 2^N \to \mathbb{R}$ is **Submodular** iff for all $A, B \subseteq N$

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$$

Equivalent Defn:

$\forall\ A \subset B \subset N, \quad \forall j \in N - B$

$$f(A \cup \{j\}) - f(A) \geq f(B \cup \{j\}) - f(B)$$

(economy of scale).

<u>equivalent defn:</u>

$\forall A \subset N, \ \forall i, j \in N - A$

$$f(A \cup \{j\}) - f(A) \geq f(A \cup \{i\} \cup \{j\}) - f(A \cup \{i\})$$

(sometimes called local submodularity)

Notice that this means

if ~~for~~ $f(A) = g(|A|)$ for all $A \subseteq N$, $g : N \to R$

then

f is submodular $\Leftrightarrow$ g is concave.

[Easy reasoning about derivatives]

Example of a submodular fn

$G = (V, A)$ directed graph, capacities $c(A)$

write

the cut $\delta^+(S)$, $S \subseteq V$ is

$$\delta^+(S) = \{ a = (i,j) \in A ;\ i \in S ,\ j \notin S \}$$

the global cut function:

$f : 2^V \rightarrow R$ is

$$f(S) = c(\delta^+(S)) = \sum_{a \in \delta^+(S)} c(a)$$

defined for $S \subseteq V$

~~the s~~ given ~~$s,t \in V$~~ $s \in V$, $t \in V$, $s \neq t$, s-t cut fn:

is

$$f_{st}(S) = f(S \cup \{s\}) \quad \text{for } \forall S \subseteq V - \{s,t\}$$

f is submodular

δ_{st}

proof

$$f(A) + f(B) - f(A \cup B) - f(A \cap B)$$

$$= \sum_{\substack{i \in S \\ j \in T}} c(i,j) + \sum_{\substack{i \in T \\ j \in S}} c(i-j)$$

$\geqslant 0$ because c is non-neg.

Further examples of submodular fns.

Mutual information

Random variables $X_1 \dots X_n$

$$F(A) = I(X_A ; X_{V-A})$$

(ie. mutual information between X's in A and those outside)

this is useful in learning to partition random variables – for example, we could split $X_1 \dots X_n$ into two sets

$A \quad | \quad V-A$

which are "as independent as possible"

choose split as

$$A^* = \arg\min I(X_A ; X_{V-A})$$

$$\text{s.t.} \quad 0 < |A| < n$$

$$F(A) = I(X_A ; X_{V-A})$$

is submodular

$$F(A) = H(X_{V-A}) - H(X_{V-A} | X_A)$$

check $F(A \cup \{s\}) - F(A)$

$$F(A \cup \{s\}) = H(X_{V-A-\{s\}}) - H(X_{V-A-\{s\}} | X_{A \cup \{s\}})$$

notice

$$H(X_{V-A} | X_A) = H(X_{V-A-\{s\}} | X_{A \cup \{s\}}) + H(X_{\{s\}} | X_A)$$

so

$$F(A \cup \{s\}) - F(A) = H(X_{\{s\}} | X_A) - \left[H(X_{V-A}) - H(X_{V-A-\{s\}}) \right]$$

<u>but</u>:

$$H(X_{V-A}) - H(X_{V-A-\{s\}})$$

$$= H(X_s \mid X_{V-A-\{s\}})$$

so $F(A \cup \{s\}) - F(A) =$

$$H(X_{\{s\}} \mid X_A) - H(X_{\{s\}} \mid X_{V-A-\{s\}})$$

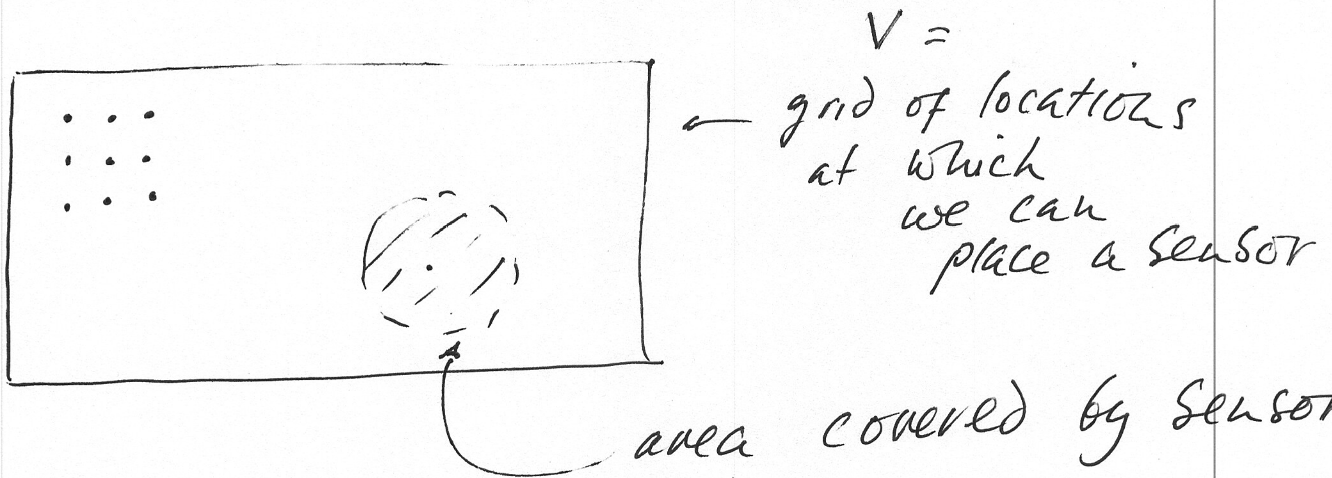
this does not <u>decrease</u> as A gets bigger

This does not increase as A gets bigger

so $\delta_s(A)$ is monotonically non-increasing.

so F is submodular

Set cover for a floor plan



$A = \{$ locations of sensors $\} \subset V$

$F_s(A) =$ area covered by sensors placed at A

equivalent formal problem.

W a finite set

S_i n subsets, $S_i \subseteq W$

$A \subseteq \{1, \ldots n\}$ $\qquad F_f(A) = \left| \bigcup_{i \in A} S_i \right|$

4e

$F_s(A)$ is submodular.

consider $A \subseteq B$

$$F_s(A \cup \{s\}) - F_s(A) \geq F_s(B \cup \{s\}) - F_s(B)$$

total area covered by sensors in $A \cup \{s\}$

i.e: when we add a sensor to a big set, the increase in coverage isn't bigger than when we add to a small set.

Similar argument applies to $F_f(A)$

Important facts:

1) Submodular function minimization is polynomial (though not spectacularly easy or efficient, $O(n^8 \log^2 n)$ or worse).

2) Submodular function max is hard, but greedy alg does very well.

thm: given F monotonic, submodular, $F(\phi) = 0$, the greedy alg gives A_{greedy} such that

$$F(A_{greedy}) \geq (1 - 1/e) \max_{|A| \leq k} F(A)$$

3) given $F(A)$ submodular, symmetric (i.e. $F(A) = F(V-A)$), there is a combinatorial alg to solve

$$\operatorname*{argmin}_{A} F(A) \quad \text{st } 0 < |A| < n$$

Runs in time $\cancel{\#}\ O(n^3)$

Submodular polyhedra

. finite set N, $f: 2^N \to R$.

$$P(f) = \{ x \in \mathbb{R}^N : x(A) \leq f(A), \ \forall A \subseteq N \}$$

where $x(A) = \Sigma\{x(e) : e \in A\}$

Model: . for any $A \subseteq N$ there is an indicator vector in $[0,1]^N$

. 1 if that el. is in, 0 otherwise

. So $P(f)$ is defined by linear constraints, which are

$$I_A \cdot x \leq f(A)$$

$\uparrow$

indicator vector for A

(But there are a lot of these 2^N)

If f is submodular, $P(f)$ is a <u>submodular polyhedron</u>.

<u>Interesting linear program</u>

(for $w \geq 0$)

$$\max \quad w^T x$$
$$\text{st} \quad x \in P(f)$$

submodular

This should look hard (constraints) but it isn't.

Greedy algorithm for this case.

1: sort the elements of N so that
$w(e_1) \geq w(e_2) \geq w(e_3) \geq w(e_4) \ldots$

2: let $V_0 = \phi$

for $i = 1$ to n

$V_i = V_{i-1} \cup \{e_i\}$

$x_i^* = x^*(e_i) = f(V_i) - f(V_{i-1})$

3: report x^*

<u>Thm</u>: this alg solves the LP
(iff f submodular)

<u>Notice</u>:

- easy to show that x^* is feasible; by cases.

<u>example</u>:
assume $A = \{e_e\}$

then $I_A = \{0 \ldots 1 \ldots 0)$
$\uparrow$
ℓ'th

$$I_A \cdot x^* = f(A_{\ell-1} \cup \{e_e\}) - f(A_{\ell-1})$$

<u>but</u>: ~~$f(A) =$~~

$$f(A_{\ell-1} \cup \{e_\ell\}) + \underbrace{f(A_{\ell-1} \cap \{e_\ell\})}_{=0} \leq f(A_{\ell-1}) + f(\{e_\ell\})$$

(submodular) $\quad = 0$ because set is empty.

$$\therefore \quad f(A_\ell \cup \{e_\ell\}) - f(A_{\ell-1}) \leq f(\{e_\ell\})$$

so feasible.

8a

Notice that this problem has important properties

consider $g(w) = \left\{ \begin{array}{l} \text{value of the problem} \\ \max \quad w^T x \\ \text{st} \quad x \in P(F) \end{array} \right\}$

submod

1) $g(w)$ is a <u>convex function</u>

2) write w^A for the indicator function of A (i.e. entries in w^A csp to elements in A are 1, others are 0)

then

$$g(w^A) = F(A)$$

same F

to show this;

- order of elements does not matter,

so rearrange so that ~~us~~ $A = \{e_1 \cdots e_K\}$

$$\text{So} \quad \begin{array}{c} w^A = \{1 \;\cdots\; 1, 0 \;\cdots\; 0\} \\ \quad\;\;\; \uparrow_1 \quad\;\; \uparrow_K \quad\;\;\; \uparrow_n \end{array}$$

$W^{A^T} x^*$ in this case is

$$\sum_{i=1}^{K} \left[F(\{e_1 \cdots e_i\}) - F(\{e_1 \cdots e_{i-1}\}) \right]$$

and terms cancel to get

$$F(\{e_1 \cdots e_K\}) - F(\emptyset) = F(A).$$

8c

Now this means that

$$\min_{A \in 2^{|V|}} F(A)$$

submodular

has become

$$\min_{w \in [0,1]^N} g(w)$$

convex

if we can show that the extremal value here is integer, then we have Polynomial alg in principle.

(8d)

Maximizing submodular fn's

F monotonic, submodular

V finite set

want

$A^* \subseteq V$ such that

$$A^* = \underset{|A| \leq k}{\operatorname{argmax}} F(A)$$

(NP-hard)

Greedy alg

$A_\emptyset = \emptyset$

For $i = 1 . \; k$

$$s_i = \underset{s}{\operatorname{argmax}} F(A_{i-1} \cup s) - F(A_{i-1})$$

$$A_i = A_{i-1} \cup s$$

Application: Correlated label prediction

- problem: exploit label correlations in prediction w/o elaborate models.

have data x_i, S_i ← a set of labels.

Similarity $K(x_i, x_j)$ ← big if x_i sim to x_j, small otherwise.

want to label x_t

- Single label

$$z_{t,j} \leq \sum_{i \in \text{examples}} K(x_t, x_i)\, I(j \in S_i)$$

↑ score of assigning j to t

inequality, because x_i may propagate only some labels

- extend to sets of labels

$$s_t(S) \leq \sum_{i \in \text{examples}} K(x_t, x_i) \cdot I(S \cap S_i \neq \emptyset)$$

indicator fn.

- the confidence of a set S should be better than or equal to conf of individual labels

$$\sum_{j \in \text{labels}} z_{t,j} \, I(j \in S) \leq S_t(S)$$

<u>combine:</u>

$$\sum_{j \in \text{labels}} z_{t,j} \, I(j \in S) \leq \sum_{i \in \text{examples}} K(x_t, x_i) \, I(S \cap S_i \neq 0)$$

linear set function

set function.

it turns out that $\sum_{i \in \text{examples}} K(x_t, x_i)\, I(S \cap S_i \neq \emptyset)$ (11)

is submodular.

now, if we ~~want~~ assume some weights for class labels α, we can ask for

$$\max_{z} \; \alpha^T z$$

$$\text{st} \quad z(S) \leq f(S)$$

the submodular fn.

Attractive feats:

- Depends only on ordering of α, not values
- No model of label correlations
- works quite well.