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:

$f : 2^V \to R$ 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

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

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

≥ 0 because c is non-neg.

Submodular polyhedra

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

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

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

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

↑ 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^*

Thm: this alg solves the LP (iff f submodular)

Notice:

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

example: assume $A = \{e_e\}$

then $I_A = \{0 \cdots 1 \cdots 0)$, with the 1 at the ℓ'th position

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

but: ~~$f(A) =$~~

$$f(A_{\ell-1} \cup \{e_\ell\}) + \underbrace{\cancel{f(A_{\ell-1} \cap \{e_\ell\})}}_{=0 \text{ because set is empty.}} \leq f(A_{\ell-1}) + f(\{e_\ell\})$$

(submodular)

$$\therefore \quad f(A_\ell \cup \{e_\ell\}) - f(A_{\ell-1}) \leq f(\{e_\ell\}) \qquad \text{so feasible.}$$

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

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

indicator fn.

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

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

<u>combine</u>:

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

linear set function

set function.

it turns out that $\sum_{i \in \text{examples}} K(x_t, x_i)\, I(S \cap S_i \neq 0)$ (11)

is <u>submodular</u>.

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.

<u>Attractive feats</u>:

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