Lecture # 23

Submodular functions and discrete optimization

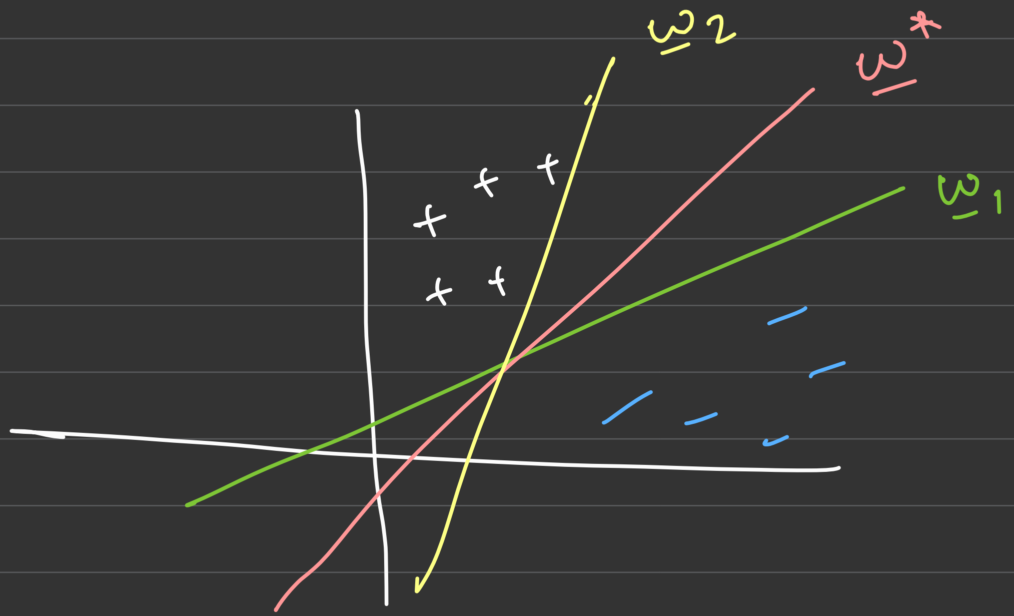
E1 260

- Combinatorial optimization in ML
- Submodular functions
- maximizing monotone submodular function
  - Greedy method
  - $\left(1-\frac{1}{e}\right)$ approximation

TA session (HW 3) : 22nd monday 1800 hrs.

— Francis Bach (monograph)

So far, we have seen convex optimization problems in ML



Classify '+' and '−' by finding a seperating hyperplane

Solve for the best vector the minimizes the

$$L(\underline{w}) \propto \frac{1}{\text{size of margin}}$$

$$\underline{w}^* = \arg\min_{\underline{w}} L(\underline{w})$$

Feature Selection:

- Predict Y from a subset $X_A = \{X_{i_1}, \ldots X_{i_k}\}$
- Given random variables $Y, X_1, \ldots, X_n$

$|A| = 2 = k$

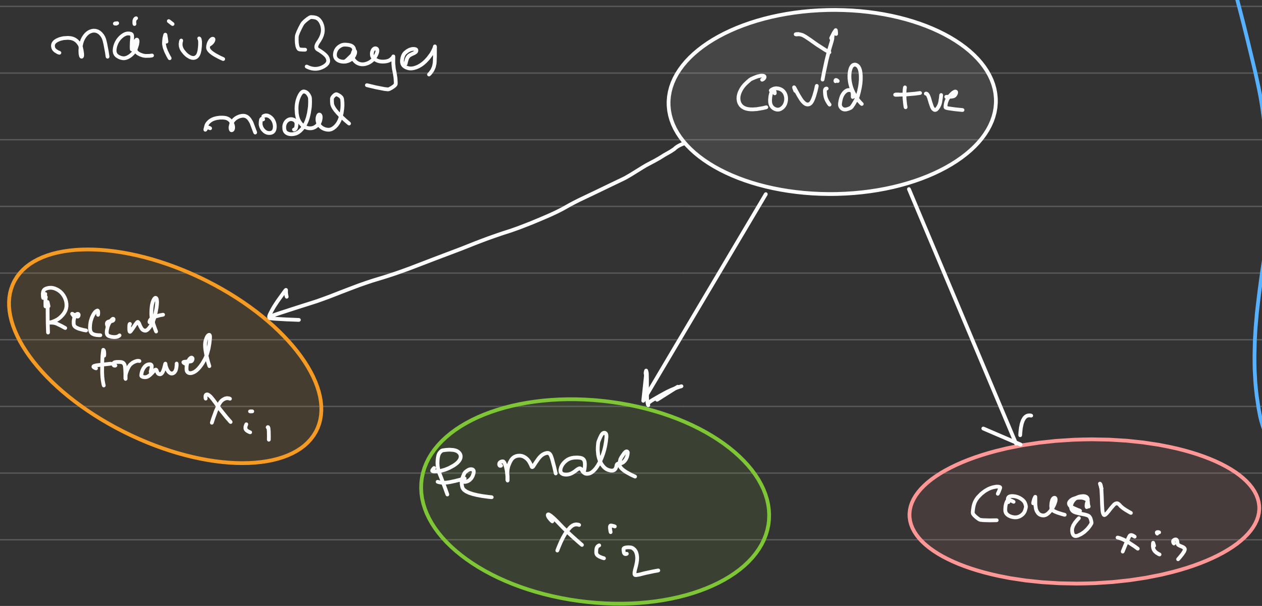
$A_1 = \{x_1, x_3\}$

$A_2 = \{x_1, x_4\}$.

$\vdots$

$A_{500} = \{x_5, x_6\}$

naive Bayes model



Wish to select K most informative features:

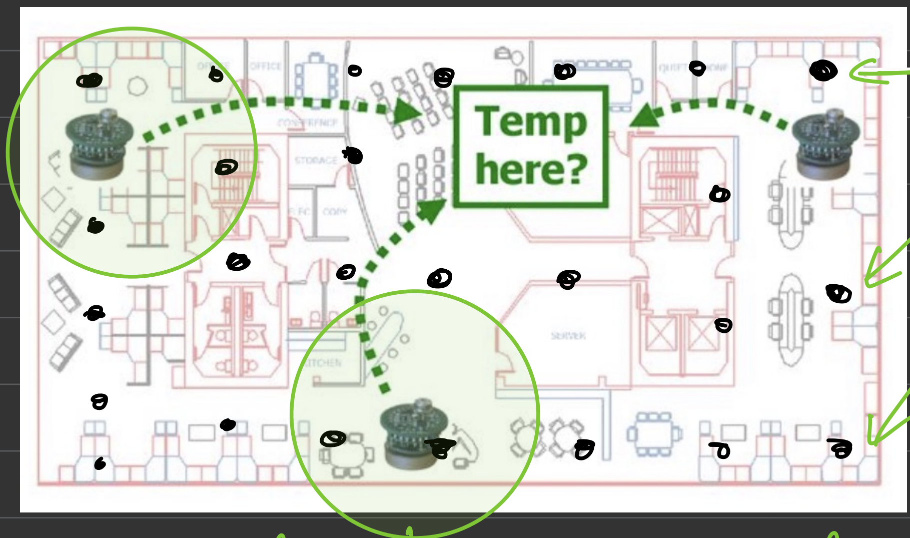
$$A^* = \arg\max \, IG(X_A ; Y) \quad \text{s.t. } |A| \le k$$

Information gain:

$$IG(x_A ; Y) = \underbrace{H(Y)}_{\substack{\text{uncertainty before} \\ \text{knowing } x_A}} - \underbrace{H(Y \mid x_A)}_{\substack{\text{uncertainity after} \\ \text{knowing } x_A}}$$

This is a combinatorial problem !!

Sensor placement (set cover problem)



Possible locations

How to place K sensors out of V candidate positions to increase the coverage?

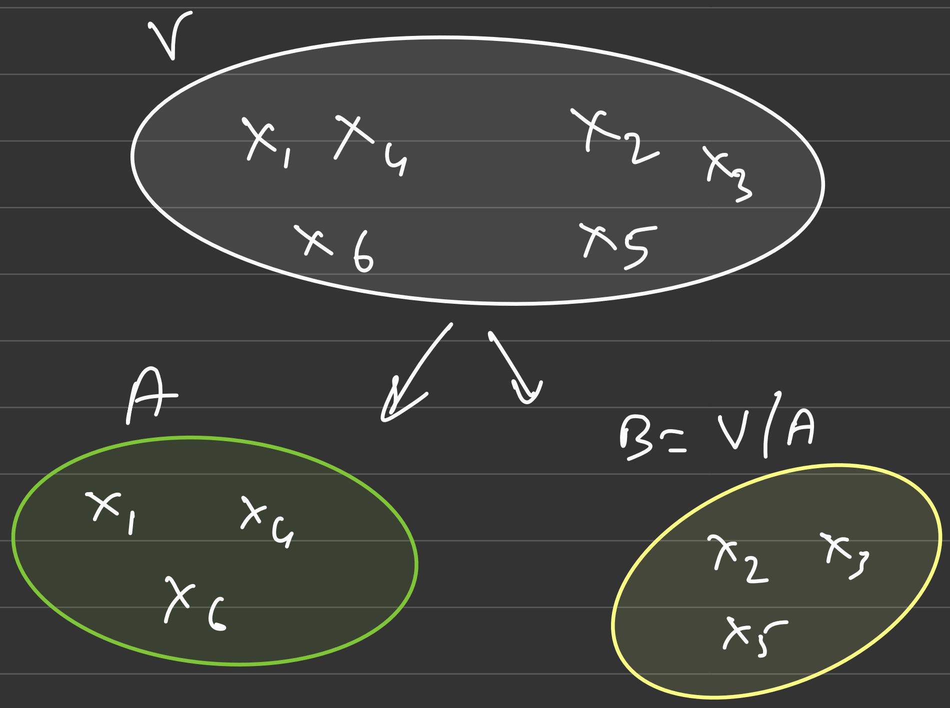
Nodes predicts/measures values with some radius/coverage.

Factoring distributions:

Given random variables $x_1, \ldots, x_N$, partition V them into set A and $B = V / A$ that are as independent as possible

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

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



$$I(X_A ; X_{V/A})$$
$$= H(X_B) - H(X_B \mid X_A)$$

Again, combinatorial !!

Set functions:

$$f : 2^X \rightarrow \mathbb{R}$$

→ Takes as input a set ; inputs are subsets of the ground set $X = \{1, 2, \ldots, N)$

→ 2^X is the power set (set of all subsets)

minimization (lly maximization) of a set function

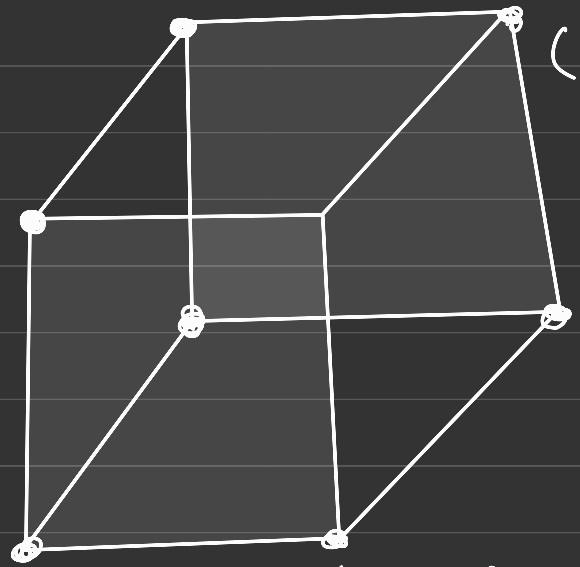
$$\min_{A \subset X} F(A) \equiv \min_{A \in 2^X} F(A)$$

s.t. constraints on the subset A

Reformulation as Boolean function:

$$\min_{\underline{\omega} \in \{0,1\}^N} f(\underline{\omega}) \quad \text{with} \quad \forall A \subset V$$

$$f(1_A) = F(A)$$



$(1,1,1) \sim \{1,2,3\}$

$(1,1,0) \sim \{1,2\}$

$(0,1,0) \sim \{2\}$

$(0,0,0) \sim \{\ \}$

$$f(\underline{w}) = \sum_{i=1}^{N} w_i I_i$$

$$(P) \quad \underset{\underline{w} \in \{0,1\}^N}{\text{maximize}} \quad f(\underline{w})$$

$$\text{s.t.} \quad \underbrace{\|\underline{w}\|_0}_{nnz(\underline{w}) = k} \leq k$$

Optimally solve this we have to exhaustively enumerate over all k-sparse vectors

$f(w)$: concave function

<u>Convex relaxation:</u>

$$(P_c) \quad \text{maximize} \quad f(w)$$

$$\text{s.to} \quad \underline{w} \in [0,1]^N \quad \text{: box constraint}$$

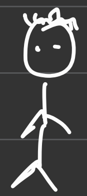
$$0 \leq w_i \leq 1$$

$$\|\underline{w}\|_1 \leq k$$

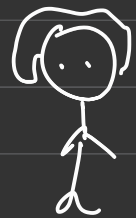
(best convex approximation of l_0-norm)

Key property : "Diminishing returns"

A



B



Bank A/C

Balance : 1,00,000

Balance : 100

Bank A/C

Cash back (PayTM)

+ 50 Rs

+ 50 Rs

Submodular functions:

A set function is said to be submodular if and only if

$$f(B \cup \{i\}) - f(B) \leq f(A \cup \{i\}) - f(A)$$

$\forall \; A \subseteq B \subseteq X$ and $i \notin B$

Equivalent definition:

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

$\forall A, B \subseteq X$

- Equality leads to modular functions
- $f(\phi) = 0$

To show equivalence:

Let $A' = A \cup \{i\}$ and $B' = B$

$$f(A \cup \{i\}) + f(B)$$
$$= f(A') + f(B') \geq f(A' \cap B') + f(A' \cup B')$$
$$= f(A \cup \{i\} \cap B) + f(A \cup \{i\} \cup B)$$
$$= f(A) + f(B \cup \{i\})$$

Super modular: f is super modular if and only if $-f$ is submodular.

Submodularity in ML:

minimization of submodular functions.

Lovasz Extension + duality

clustering, structure learning MAP inference in Markov random fields

maximization of submodular functions:

difference of submodular

Active learning, feature selection, ranking

$$\min_{x \in \mathcal{X}} \quad f(x) - g(x)$$

Difference convex fn.

$f(x)$ & $g(x)$ are convex

SCP :

$$\min_{x \in \mathcal{X}} \quad f(x) - g(x_t) - \nabla g(x_t)^T x_0$$

Example of Submodular function:

E.g.: flows, set cover, differential entropies

Entropy:

Given p random variables $X_1, \dots X_p$

$F(A)$ as the joint entropy of variables $(X_k)_{k \in A}$

$\Rightarrow$ $F(A)$ is submodular

if $A \subseteq B$ and $k \notin B$

$$F(A \cup \{k\}) - F(A) = H(X_A, X_k) - H(X_A)$$

$$= H(X_k \mid X_A)$$

[conditioning reduces Entropy) $\geq H(X_k \mid X_B)$

$$= F(B \cup \{k\}) - F(B)$$

Maximizing Submodular functions:

$$\text{maximize} \quad f(A)$$
$$\text{s. to} \quad |A| \leq K$$
$$A \subseteq V$$

Nemhauser (1978):

If f is submodular, monotone increasing, and nonempty

$$f(A \cup \{i\}) \geq f(A) \qquad f(\phi) = 0$$

Then Greedy algorithm: $A = \phi$

for $i = 1, 2, \ldots K$

$$i \leftarrow \underset{i \notin A}{\arg\max} \left[f(A \cup \{i\}) - f(A) \right]$$

$A \leftarrow A \cup \{i\}$; return A

The above greedy method satisfies :

$$f(A) \geq \underbrace{\left(1 - \frac{1}{e}\right)}_{63\%} f(A_{opt})$$

$$f(A_{opt}) = \arg\max_{A \subseteq V \,;\, |A| = k} f(A) \qquad \left[\text{Exhaustive search}\right]$$

⟹ Although this bound is not that tight, results are close to exhaustive search in practice (whenever, verifiable).

Claim: pick any $A \subseteq V$ such that $|A| < k$.

Then

$$\max_{i \in V} \left[f(A \cup \{i\}) - f(A) \right] \geq \frac{1}{k} \left[f(A_{opt}) - f(A) \right]$$

Proof:

Let $A_{opt} \setminus A = \{ i_1, \ldots, i_p \}$ so that $p \leq k$

Then $f(A_{opt}) \leq f(A_{opt} \cup A)$ (monotonicity)

$$= f(A) + \sum_{j=1}^{p} f(A \cup \{i_1, \ldots, i_j\}) - f(A \cup \{i_1, \ldots, i_{j-1}\})$$

(submodularity) $\leq f(A) + \sum_{j=1}^{p} \left[f(A \cup \{i_j\}) - f(A) \right]$

$$\leq f(A) + \sum_{j=1}^{p} \max_{i \in A} \left[f(A \cup \{i\}) - f(A) \right]$$

$(p \leq k) \quad \leq f(A) + k \max_{i - A} \left[f(A \cup \{i\}) - f(A) \right]$

approximation proof:

Let A^k be the solution of the greedy method at step K. Then from the previous result

$$\Rightarrow f(A^K) - f(A^{K-1}) \geq \frac{1}{K} \left[f(A^{opt}) - f(A^{K-1}) \right]$$

$$f(A^K) \geq \frac{1}{K} f(A_{opt}) + \left(1 - \frac{1}{K}\right) f(A^{K-1})$$

$$\geq \frac{1}{K} f(A_{opt}) + \left(1 - \frac{1}{K}\right) \left[\frac{1}{K} f(A_{opt}) + \left(1 - \frac{1}{K}\right) f(A^{K-2}) \right]$$

$(p \leq k) \qquad \leq f(A) + k \max_{i - A} \left[f(A \cup \{i\}) - f(A) \right]$

<u>approximation proof:</u>

Let A^i be the solution of the greedy method at step i. Then from the previous result

$$\Rightarrow \quad f(A^i) - f(A^{i-1}) \geq \frac{1}{K} \left[f(A^{opt}) - f(A^{i-1}) \right]$$

$$f(A^{opt}) - f(A^i) \leq \left(1 - \frac{1}{K}\right) \left[f(A^{opt}) - f(A^{i-1}) \right]$$

Combining for every iteration : $1 \leq i \leq K$

$$f(A^{opt}) - f(A^K) \leq \left(1 - \frac{1}{K}\right)^K \left[f(A^{opt}) - \underbrace{f(\phi)}_{\geq 0} \right]$$

$$\Rightarrow \quad f(A^K) \geq f(A^{opt}) - \left(1 - \frac{1}{K}\right)^K \left[f(A^{opt}) \right]$$

Using the fact that $1 - x \leq e^{-x}$

$$\Rightarrow \left(1 - \frac{1}{K}\right)^K \leq \frac{1}{e}$$

$$\Rightarrow \quad f(A^K) \geq \left(1 - \frac{1}{e}\right) f(A^{opt})$$

$\blacksquare$