Problem background:

Take smartphone market as an example, there is different models of smartphone （$D = \{r_1, r_2, \dots, r_n\}$） . Each model has different price, pixel, battery capacity, cooling capacity and etc （$r_i = [r_i^1, r_i^2, \dots, r_i^d]$） .

Now brand P has m models ($P = \{p_1, p_2, \dots, p_m\} \subset D$), and have user data set $W = \{w_1, w_2, \dots, w_x\}$. Each $w_i = [w_i^1, w_i^2, \dots, w_i^d]$ present the preference weight vector of a single user.

The score of a model r_i respecting to a user w_i is the dot product $r_i \cdot w_i$. A model covers a user when the score of it ranks top k among D.

Q: How to introduce a new product p such that maximizes the cover ratio $cp(p, P, k) = \frac{|\{w | \forall w \in W, \{P \cup \{p\}\} \cap TopK(w) \neq \emptyset\}|}{|W|}$ and the product creation cost $C(p)$ is no greater than creation budget B?

Problem Definition:

Input:

global model dataset $D = \{r_1, r_2, \dots, r_m\}$, $r_i = [r_i^1, r_i^2, \dots, r_i^d]$,

Another model dataset $P = \{p_1, p_2, \dots, p_n\} \subset D$

Model creation budget B,

a positive integer k,

a creation cost function $C(p_i) = \Sigma c_j p_j$,

a user preference tuple set $W = \{w_1, w_2, \dots, w_x\}, w_i = [w_i^1, w_i^2, \dots, w_i^d]$,

The score of a model r_i respecting to a user w_i is the dot product $r_i \cdot w_i$

A model covers a user when the score of it ranks top k among D.

Output:

find a new product p such that satisfies the constrain $C(p) \leq B$

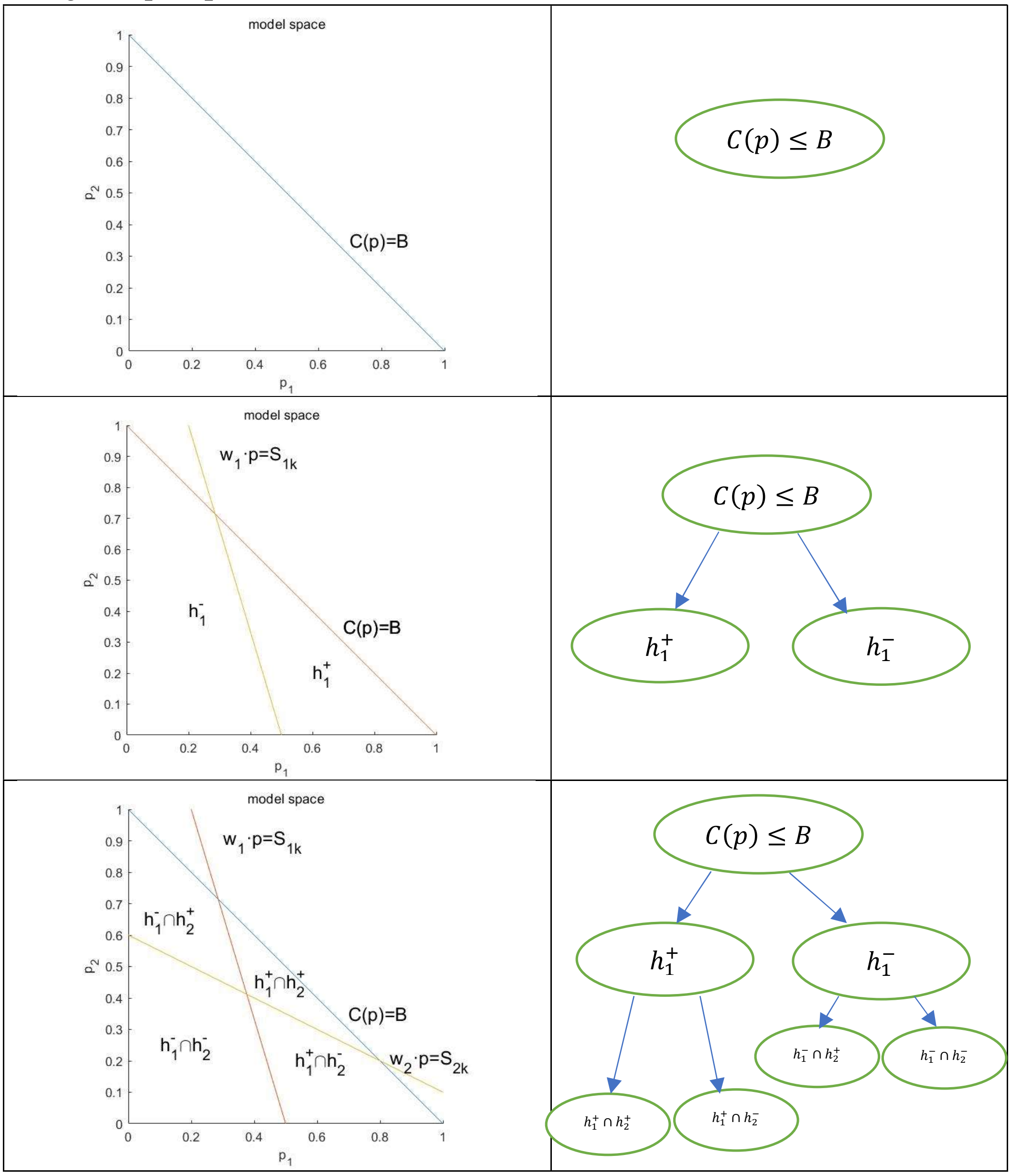
and maximize $P \cup \{p\}$'s cover ratio $cp(p, P, k) = \frac{|\{w | \forall w \in W, \{P \cup \{p\}\} \cap TopK(w) \neq \emptyset\}|}{|W|}$

Baseline solution example

Take $d = 2$ as example,

where S_{ik} is the score to rank top k respecting to w_i

1. In model space, only the half-space C(p)≤B is valid
2. If we want to cover w_1, the score of new model p should greater than S_{1k}
   The line $w_1 \cdot p = S_{1k}$ divides the space $C(p) \leq B$ into 2 part.
   One is h_1^+, where any model in it could cover w_1;
   Another is h_1^-, where any model in it couldn't cover w_1.
3. If we want to cover w_2, the score of new model p should greater than S_{2k}
   The line $w_2 \cdot p = S_{1k}$ divides the space h_1^+ into 2 part.
   One is $h_1^+ \cap h_2^+$, where any model in it could cover w_1, w_2;
   Another is $h_1^+ \cap h_2^-$, where any model in it could only cover w_1.
   The line $w_2 \cdot p = S_{2k}$ divides the space h_1^- into 2 part.
   One is $h_1^- \cap h_2^+$, where any model in it could only cover w_2;
   Another is $h_1^- \cap h_2^-$, where any model in it couldn't only any user.
4. Return region $h_1^+ \cap h_2^+$



Baseline solution:

1. For each w_i in W, based on $w_i \cdot p \leq S_{ik}$ using cell tree divide model space $C(p) \leq B$, where S_{ik} is the score to rank top k respecting to w_i
2. count each divided region's user cover number
3. Return the region that satisfies $C(p) \leq B$ and with the greatest cover count

Advance solution example:

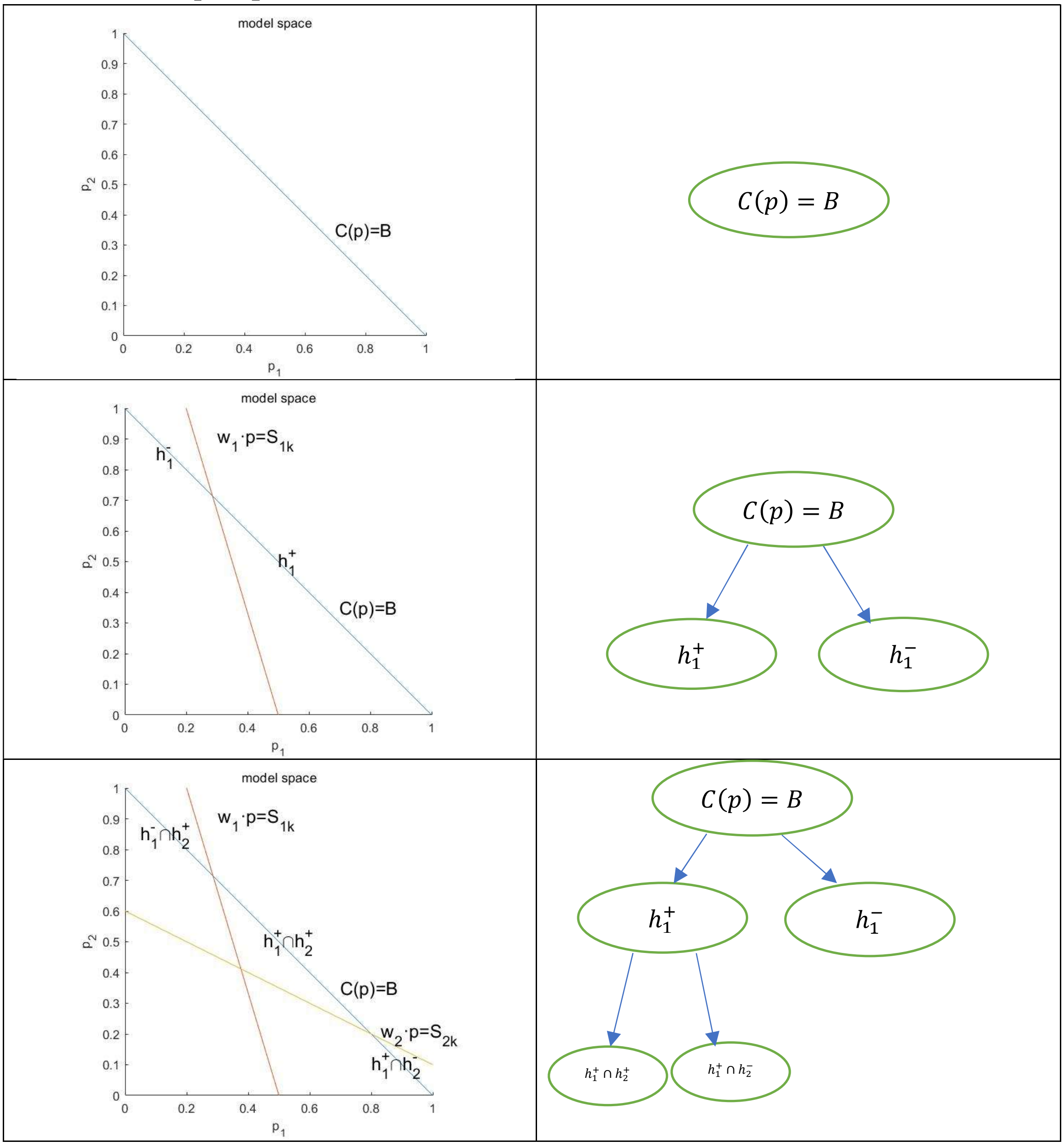
Take $d = 2$ as example,

where S_{ik} is the score to rank top k respecting to w_i

1. discard all weight vectors such that have been cover by P:

$$W = W - \{w | \forall w \in W, P \cap TopK(w) \neq \emptyset\}$$

2. In model space, only consider the space $C(p) = B$
3. The line $w_1 \cdot p = S_{1k}$ divides the line $C(p) = B$ into 2 part.
   One is h_1^+, where could cover w_1;Another is h_1^-, where couldn't cover w_1.
4. The line $w_2 \cdot p = S_{2k}$ divides the space h_1^+ into 2 part.
   One is $h_1^+ \cap h_2^+$, could cover w_1, w_2; Another is $h_1^+ \cap h_2^-$, where could only cover w_1.
5. Return line $h_1^+ \cap h_2^+$



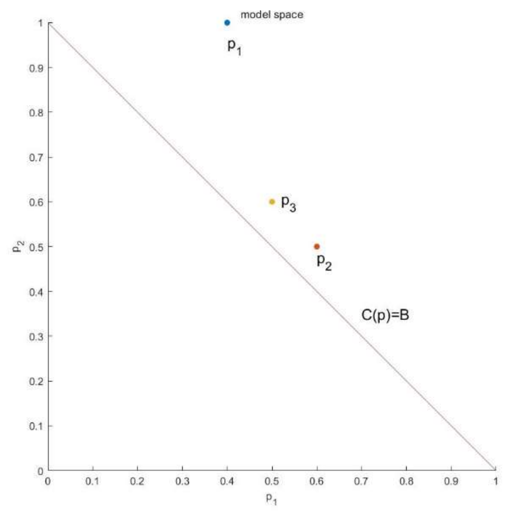
Advance solution:

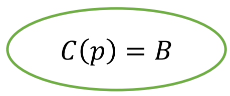
1. Preprocessing I, discard all user that P already covers:
$$W = W - \{w|\forall w \in W, P \cap TopK(w) \neq \emptyset\}$$
2. we only consider $C(p) = B$
3. Preprocessing II, discard all w such that $w \cdot p = S_{ik}$ doesn't intersect with $C(p) = B$
$$W = W - \{w|\forall w \in W, \{p|w \cdot p = S_k\} \cap \{p|C(p) = B\} = \emptyset\}$$
4. For each w_i in W, based on $w_i \cdot p \leq S_{ik}$ using cell tree divide model space $C(p) = B$, where S_{ik} is the score to rank top k respecting to w_i
5. count each divided region's user cover number
6. Return the region that satisfies $C(p) = B$ and with the greatest cover count

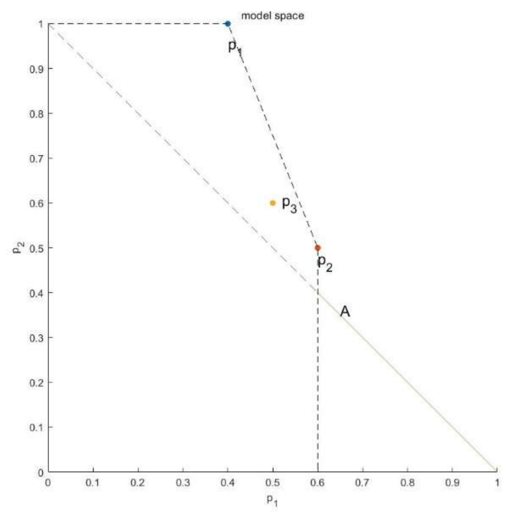
Advance solution version-2 example:

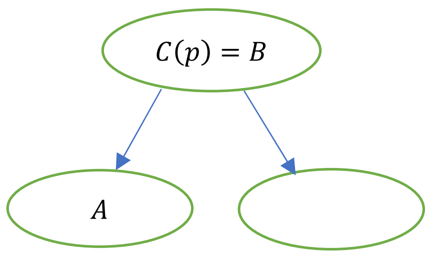
Suppose $d = 2$ and $\{p_1, p_2, p_3\} \subset P$

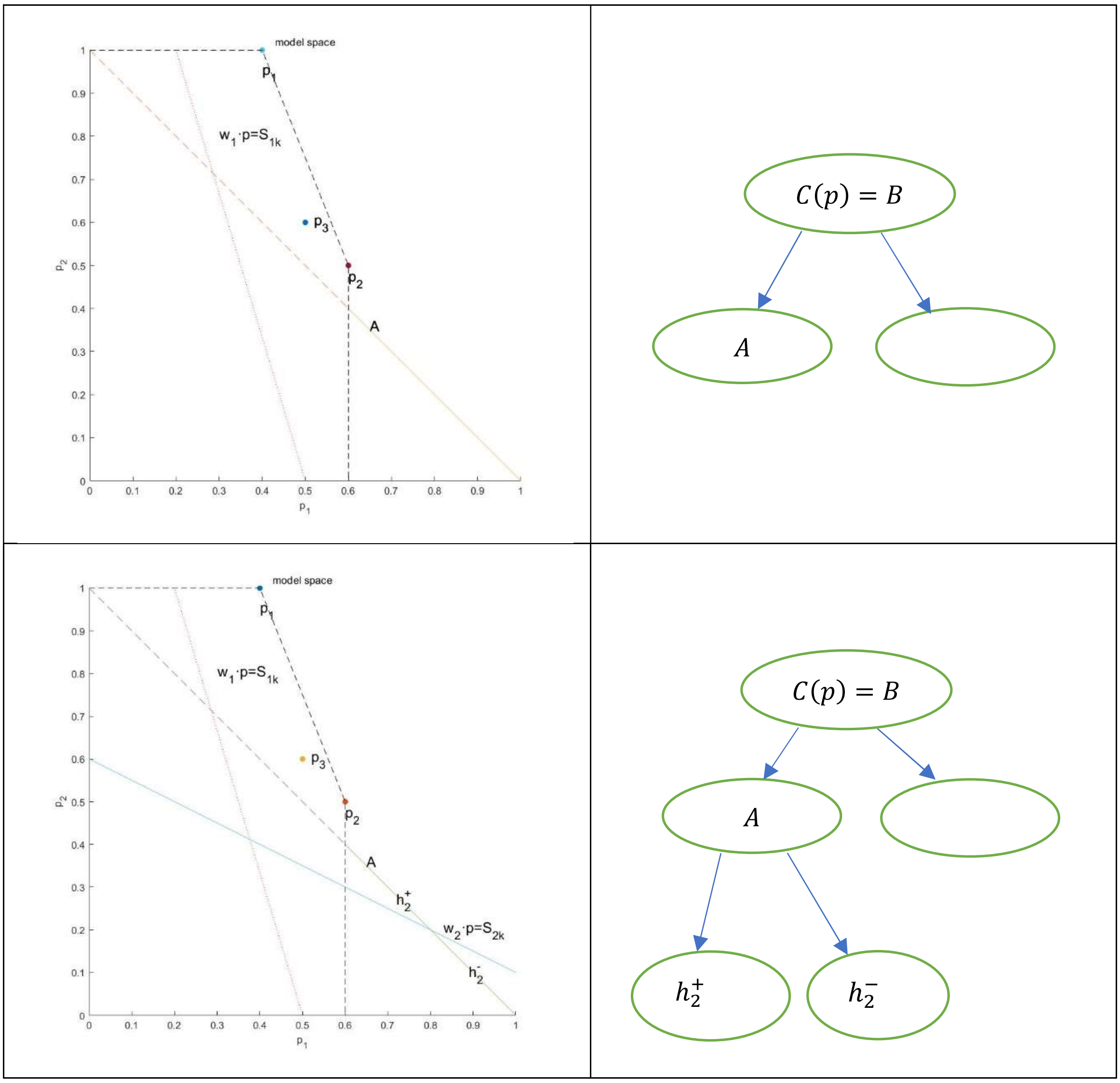
1. Preprocessing I, discard all user that P already covers:
$$W = W - \{w | \forall w \in W, P \cap TopK(w) \neq \emptyset\}$$
2. we only consider $C(p) = B$
3. Find the biggest convex hull of point set P (name it as B_CH)
4. Preprocessing II, on $C(p) = B$, discard the part which cover by convex hull (***see proof 1***)
$$A = \{p | C(p) = B\}$$
$$A = A - B\_CH$$
5. Preprocessing III, discard all w such that $w \cdot p = S_{ik}$ doesn't intersect with A
$$W = W - \{w | \forall w \in W, \{p | w \cdot p = S_k\} \cap \{p | A\} = \emptyset\}$$
As shown below, we only need the solid part of $C(p) = B$
Because $w_1 \cdot p = S_{1k}$ doesn't intersect with A, we discard and don't need to consider w_1
6. Add cover count for each piece of A based on $w \cdot p = S_k$ intersect with $C(p) = B$ but not A
Here all region of A plus 1 count because of w_1
7. The line $w_2 \cdot p = S_{2k}$ divides the line A into 2 part.
One is h_2^+, could cover w_2; Another is h_2^-, where couldn't cover w_2.
8. Return segmentation h_2^+











Advance solution version-2:

1. Preprocessing I, discard all user that P already covers:
$$W = W - \{w|\forall w \in W, P \cap TopK(w) \neq \emptyset\}$$
2. we only consider $C(p) = B$
3. Find the biggest convex hull of point set P (name it as B_CH)
4. Preprocessing II, on $C(p) = B$, discard the part which cover by convex hull (***see proof 1***)
$$A = \{p|C(p) = B\}$$
$$A = A - B\_CH$$
5. Preprocessing III, discard all w such that $w \cdot p = S_{ik}$ doesn't intersect with A
$$W = W - \{w|\forall w \in W, \{p|w \cdot p = S_k\} \cap \{p|A\} = \emptyset\}$$
6. Add cover count for each piece of A based on $w \cdot p = S_k$ intersect with $C(p) = B$ but not A
7. For each w_i in W, based on $w_i \cdot p \leq S_{ik}$ using cell tree divide model space A, where S_{ik} is the score to rank top k respecting to w_i
8. count each divided region's user cover number
9. Return the region that with the greatest cover count