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 $C(p_i) = \Sigma c^j p_i^j$,

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 $cp(p, P, k) = \frac{|\{w | \forall w \in W, \{P \cup \{p\}\} \cap TopK(w) \neq \emptyset\}|}{|W|}$

Hit:

The main difference between baseline and advance-1 is:

1. advance-1 only consider weight vectors that not covered by P
2. advance-1 only consider $C(p) = B$ rather than $C(p) \leq B$
3. advance-1 only consider the weight vectors w that $w \cdot p = S_k$ intersects with $C(p) = B$

The main difference between advance-1 and advance-2 is:

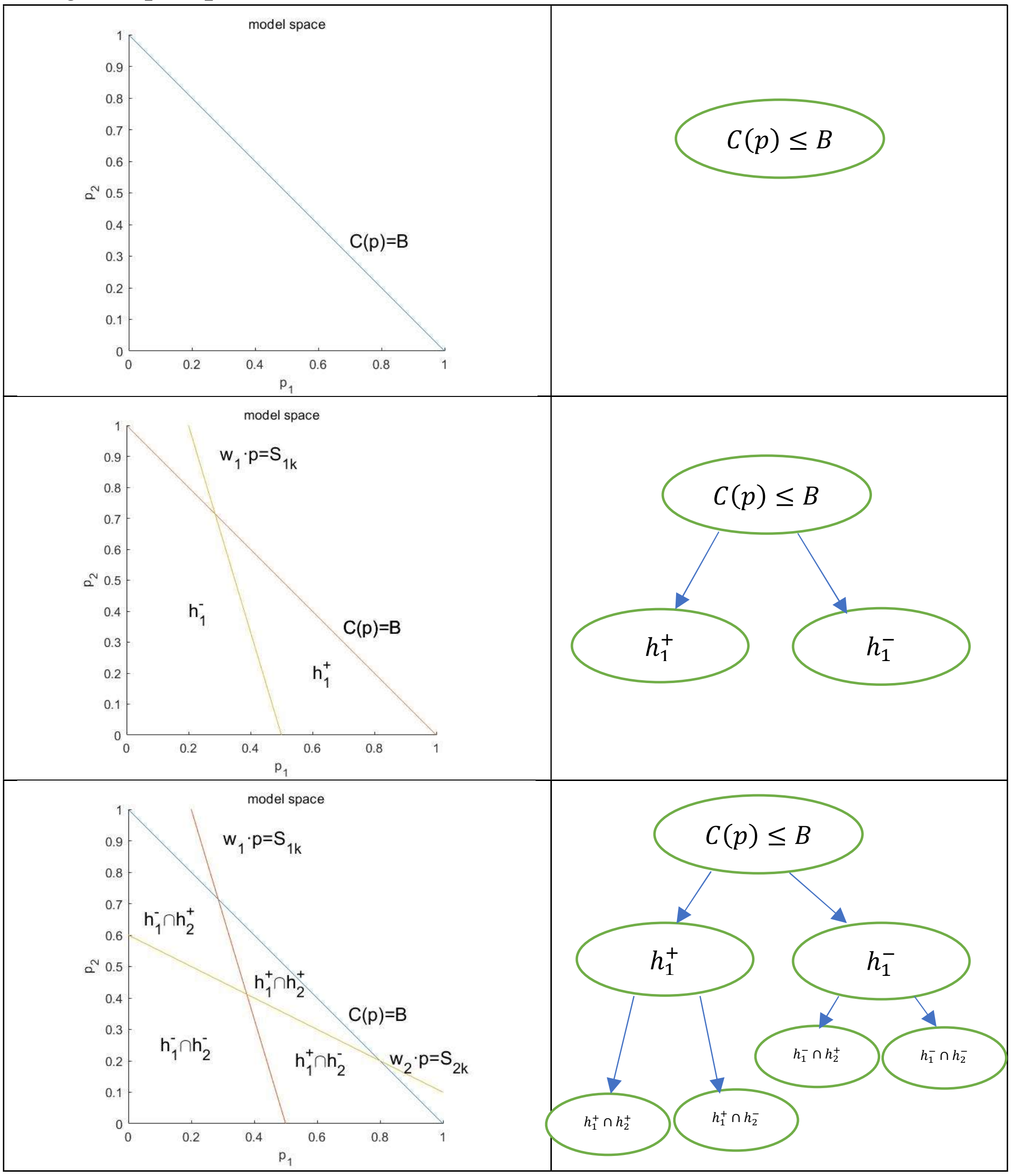
1. advance-2 considers convex hull to reduce the range of $C(p) = B$

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 the new model p should greater than S_{1k}
   The line $w_1 \cdot p = S_{1k}$ divides the space $C(p) \le B$ 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 the 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 ($\boldsymbol{kSPR}$):

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

   Time complexity: $O(n^d)$, where n is the capacity of D, d is the dimension of a model.

Advance solution version-1 example:

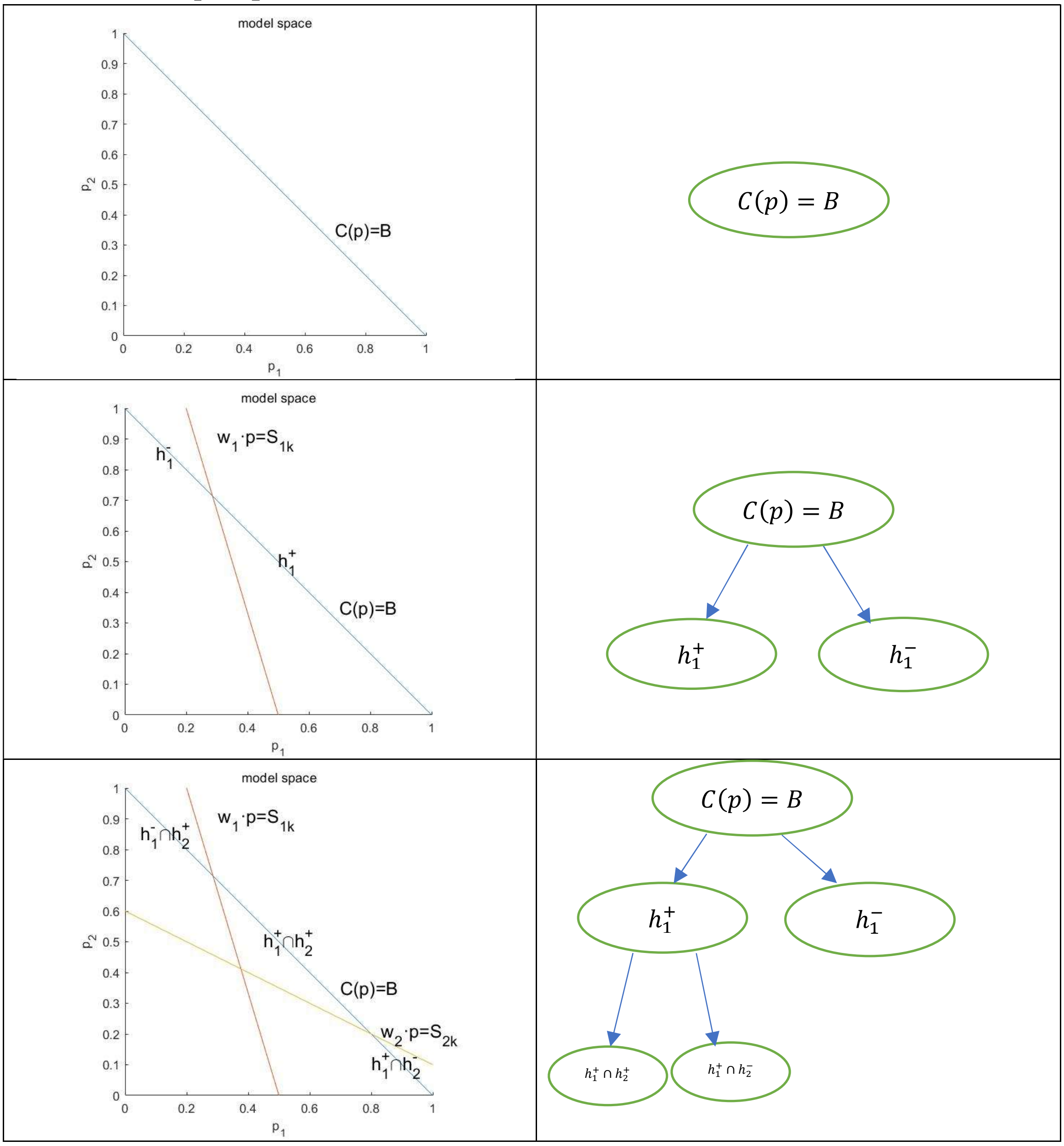
Take $d = 2$ as example,

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 version1:

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

Time complexity: $O(n_{new}^{d-1})$,
where n_{new} is the number of w such that not cover by P and $w \cdot p = S_k$ intersect with $C(p) = B$;
d is the dimension of a model.

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 is in 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 $C(p) = B$

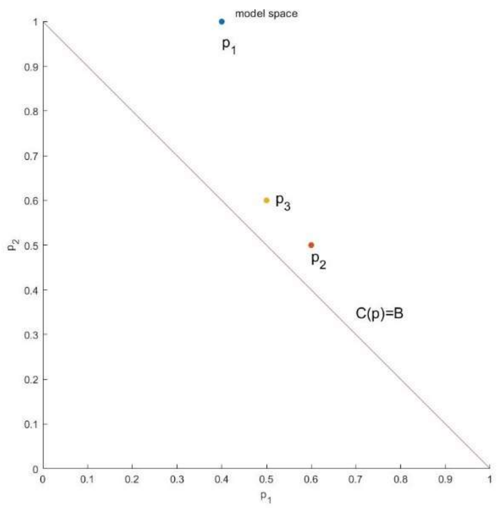
$$W = W - \{w|\forall w \in W, \{p|w \cdot p = S_k\} \cap \{p|C(p) = B\} = \emptyset\}$$

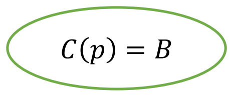
   As shown below, we only need the solid part of $C(p) = B$, which is A

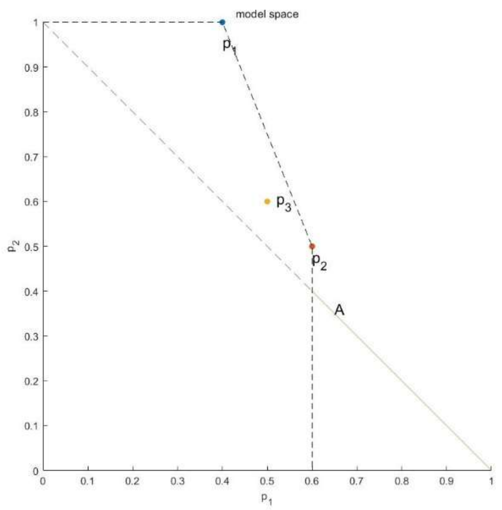
6. From line $w_1 \cdot p = S_{1k}$, we could know all of A is in h_1^+, which means A could cover w_1
7. The line $w_2 \cdot p = S_{2k}$ divides the line A into 2 part.

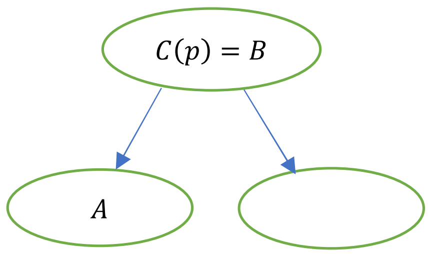
   One is $h_1^+ \cap h_2^+$, could cover w_2; Another is $h_1^+ \cap h_2^-$, where couldn't cover w_2.

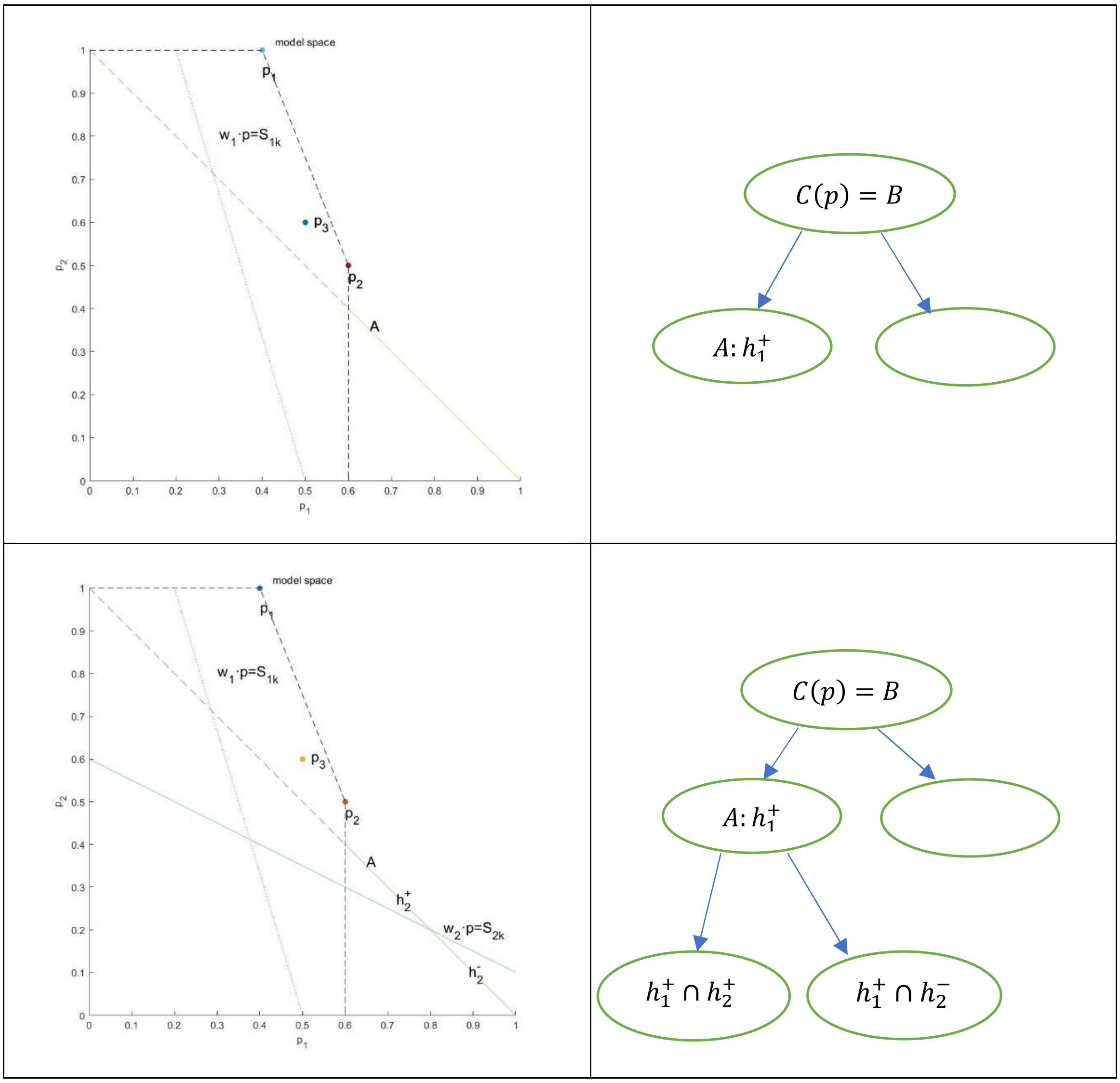
8. Return segmentation $h_1^+ \cap h_2^+$











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 is in 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 $C(p) = B$

$$W = W - \{w|\forall w \in W, \{p|w \cdot p = S_k\} \cap \{p|C(p) = B\} = \emptyset\}$$

6. 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
7. count each divided region's user cover number
8. Return the region that with the greatest cover count

Time complexity: $O(n_{new}^{d-1})$,
where n_{new} is the number of w such that not cover by P and $w \cdot p = S_k$ intersect with A;
d is the dimension of a model;
Compare with version1, version2's n_{new} is no more than version1's n_{new}.

Proof 1:
To proof:
if a model is covered by a convex hull and this model could cover weight vector w,
then one of the models that form this convex hull can also cover w.

Proof:
Suppose point O is the original point; point C is inside convex hull; D is the crossover point between OC and convex hull; on convex hull, D is the on the hyperplane $H_1H_2 \dots H_d$. Then

$$\overrightarrow{OD} = \lambda_1\overrightarrow{OH_1} + \lambda_2\overrightarrow{OH_2} + \cdots + \lambda_{d-1}\overrightarrow{OH_{d-1}} + \lambda_d\overrightarrow{OH_d}, \quad \sum_{i=1}^{d} \lambda_i = 1;\ \forall i \in [1,d], \lambda_i \geq 0$$

Next,

$$w \cdot \overrightarrow{OD} = w \cdot (\lambda_1\overrightarrow{OH_1} + \lambda_2\overrightarrow{OH_2} + \cdots + \lambda_{d-1}\overrightarrow{OH_{d-1}} + \lambda_d\overrightarrow{OH_d})$$

,
Without loss of generality, suppose $\forall i \in [1,d], w \cdot \overrightarrow{OH_1} \geq w \cdot \overrightarrow{OH_i}$

Then

$$\begin{aligned}
& w \cdot \overrightarrow{OC} \\
& < w \cdot \overrightarrow{OD} \\
& = w \cdot \left(\lambda_1\overrightarrow{OH_1} + \lambda_2\overrightarrow{OH_2} + \cdots + \lambda_{d-1}\overrightarrow{OH_{d-1}} + \lambda_d\overrightarrow{OH_d}\right) \\
& \leq w \cdot (\lambda_1\overrightarrow{OH_1} + \lambda_2\overrightarrow{OH_1} + \cdots + \lambda_{d-1}\overrightarrow{OH_1} + \lambda_d\overrightarrow{OH_1}) \\
& = w \cdot \overrightarrow{OH_1}
\end{aligned}$$

That is

$$w \cdot \overrightarrow{OC} < w \cdot \overrightarrow{OH_1}$$

Proof ends.

Proof 1's corresponding original lemma in $k - Hit\ query$:

Lemma 4. Let p be a tuple in D which is inside $Conv(D)$ but is not on the surface of $Conv(D)$. Then, $HP(p) = 0$.

Proof of Lemma 4:
Firstly, we show that for any point $p' \in D$ and any (linear) function $f \in \mathcal{L}$ associated with its weight vector w, $f(p')$ is equal to the length of the projection of the vector from the origin to point p' on vector w. This can be shown as follows. We denote the vector from the origin to point p' by $\overrightarrow{op'}$. Let θ be the angle between vector ω and vector $\overrightarrow{op'}$. $f(p') = \omega \cdot p = ||\omega||||\overrightarrow{op'}||cos\theta$. Since the norm of the weight vector ω is 1 (by definition) (i.e., $\left||\omega|\right| = 1$), we have $f(p') = ||\overrightarrow{op'}||cos\theta$. Note that $||\overrightarrow{op'}||cos\theta$ is equal to the length of the projection of vector $\overrightarrow{op'}$ on vector ω, which completes the proof.
Secondly, we show that $HP(p) = 0$. For any weight vector ω where its corresponding utility function f is in $\mathcal{L}$, there exists a point p' on the surface of $Conv(D)$ such that the length of the projection of the vector from the origin to point p' on vector ω is greater than the length of the projection of the vector from the origin to point p on vector ω. This means that $f(p') > f(p)$. In other words, for any utility function f in L, there exists a point p' on the surface of $Conv(D)$ such that $f(p') > f(p)$. This implies that $HP(p) = 0$.