CMSC 754: Spring 2026 Dave Mount

Solutions to Homework 8

Solution 1: This problem is known as the *discrete k-center problem*. Let C^* denote the optimal set of centers, and let $r^* = r^*(P,k) = \text{dist}(P, C^*)$. Let C' denote the crude approximation, and let $r' = \text{dist}(P, C')$. As observed in the problem statement, we have $r^* \le r' \le 2r^*$. As given in the problem statement, this approximation can be computed in $O(n \log n)$ time.

To construct the coreset, build a hypercube grid of side length $\varepsilon r'/(2\sqrt{d})$. The cells of this grid each have diameter $\varepsilon r'/2$, which we denote by δ. Hash all the points of P into these grid squares and select one representative point from each nonempty grid square. Under the assumption that hashing takes $O(1)$ time per point, this can be done in $O(n)$ time. We assert that the resulting set of representatives, denoted R, is the desired coreset (see Fig. 1(a)).

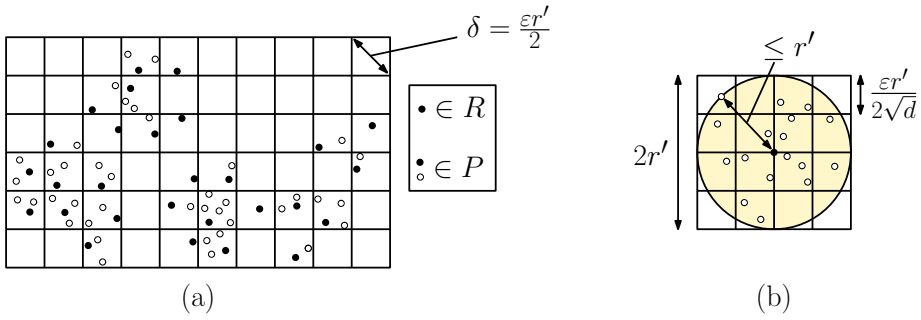


Figure 1: Construction of the k-center coreset.

We first observe that $|R| = O(k/\varepsilon^d)$. To see this, observe that every point of P lies within distance r' of one of the k points of C'. This implies that every point lies within a hypercube of side length $2r'$ centered about each of the points of C' (see Fig. 1(b)). Given our assumption that d is constant, it follows that (up to constant factors), the number of grid cells of side length $\varepsilon r'/2\sqrt{d}$ overlapping such a hypercube is at most

$$\left(\frac{2r'}{\varepsilon r'/2\sqrt{d}}\right)^d = \left(\frac{4\sqrt{d}}{\varepsilon}\right)^d = O\left(\left(\frac{1}{\varepsilon}\right)^d\right).$$

Taking the union over all the k centers of $\widehat{C}$, we have a total of at most $O(k/\varepsilon^d)$ nonempty cells, and hence $|R| = O(k/\varepsilon^d)$. The overall running time is $O(n \log n + n + |R|) = O(n \log n)$.

To establish correctness, we will show that R is an ε-coreset for the clustering problem. Let $\widehat{r}^* = r^*(R,k)$ denote the optimal covering radius for R, and let $\widehat{C}^*$ denote the set of centers for this solution. It suffices to show that $(1-\varepsilon)r^* \le \widehat{r}^* \le (1+\varepsilon)r^*$.

We first show that $r^* \le \widehat{r}^* + \delta$ by demonstrating that every point of P lies within distance $\widehat{r}^* + \delta$ of some point of $\widehat{C}^*$. (Since this distance bound holds for the particular k-element subset $\widehat{C}^* \subseteq P$, it applies to the optimal k-element subset.) To see why, consider any $p \in P$. Let p' denote the representative point from p's grid cell (see Fig. 2(a)). Because R has a covering of radius $\widehat{r}^*$,

there exists a center $c' \in \widehat{C}^*$ such that $\|p' - c'\| \leq \widehat{r}^*$. Because the diameter of the cell is δ, we have $\|p - p'\| \leq \delta$, and hence by the triangle inequality,

$$\|p - c'\| \;\leq\; \|p - p'\| + \|p' - c'\| \;\leq\; \delta + \widehat{r}^*,$$

as desired. Equivalently, $r^* - \delta \leq \widehat{r}^*$.

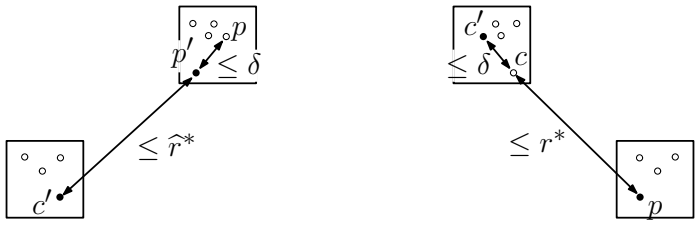


Figure 2: Correctness of the k-center coreset.

Next, we claim that $\widehat{r}^* \leq r^* + \delta$. Consider the set of centers defined as the representative points from each cell containing a point of C^*. These centers are elements of R, and hence they define a valid set of centers for R. Consider any center $c \in C^*$, and let $c' \in R$ be its representative (see Fig. 2(b)). Because the cell diameter is δ, we have $\|c - c'\| \leq \delta$. Since each point $p \in P$ (and hence each point $p \in R$) lies within distance r^* of some point of C^*, there is a center c such that $\|p - c\| \leq r^*$. Therefore, by the triangle inequality we have

$$\|p - c'\| \;\leq\; \|p - c\| + \|c - c'\| \;\leq\; r^* + \delta,$$

as desired.

Combining these bounds, we have

$$r^* - \delta \;\leq\; \widehat{r}^* \;\leq\; r^* + \delta.$$

By our choice of δ and the fact that $\widehat{r} \leq 2r^*$, we have $\delta = \varepsilon\widehat{r}/2 \leq \varepsilon r^*$. It follows that $(1-\varepsilon)r^* \leq r^* - \delta$ and $r^* + \delta \leq (1+\varepsilon)r^*$. In conclusion, we have

$$(1-\varepsilon)r^* \;\leq\; \widehat{r}^* \;\leq\; (1+\varepsilon)r^*,$$

implying that R is an ε-coreset.

Solution 2:

(a) We will show that the VC-dimension of $\Sigma = (\mathbb{R}^2, \mathcal{U})$ is 3.

VC($\mathcal{U}$) ≥ 3 : Consider the point set of Fig. 3(a), and consider the U-shaped region whose boundary passes through these points as shown. By perturbing each of the sides left/right or up/down, we can include or exclude any combination of these points. Therefore, we can generate all the subsets of $\{a, b, c\}$, implying that the VC-dimension is at least 3.

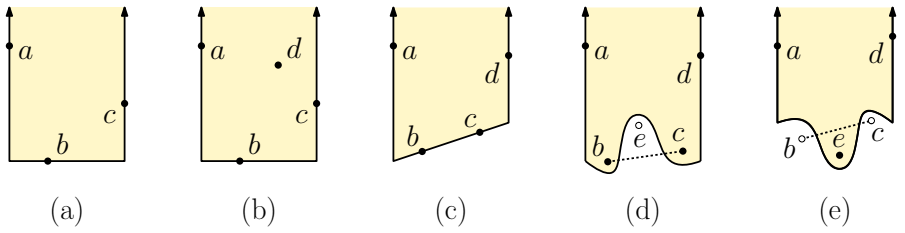


Figure 3: (a) $\text{VC}(\mathcal{U}) \geq 3$, (b) $\text{VC}(\mathcal{U}) \leq 3$, (c) $\text{VC}(\mathcal{V}) \geq 4$, and (d,e) $\text{VC}(\mathcal{V}) \leq 4$.

$\text{VC}(\mathcal{U}) \leq 3$: We claim that no 4-element point set $\{a, b, c, d\}$ can be shattered by this range space. To see this, consider four points in general position. Consider the leftmost, bottommost, and rightmost points. These constitute up to three of the points, and leaves at least one remaining point, call it d (see Fig. 3(b)). Any orthogonal U-shaped range that contains $\{a, b, c\}$ must include d, and therefore, it is not possible to generate this set.

(b) We will show that the VC-dimension of $\Sigma = (\mathbb{R}^2, \mathcal{V})$ is 4.

$\text{VC}(\mathcal{V}) \geq 4$: Consider the point set of Fig. 3(c), and consider the V-shaped region whose boundary passes through these points as shown. By perturbing the left and right sides to the left/right, we can include or exclude any combination of a and d. By raising/lower or rotating the bottom line, we can include or exclude any combination of b and c. Therefore, we can generate all of the subsets of $\{a, b, c, d\}$, implying that the VC-dimension is at least 4.

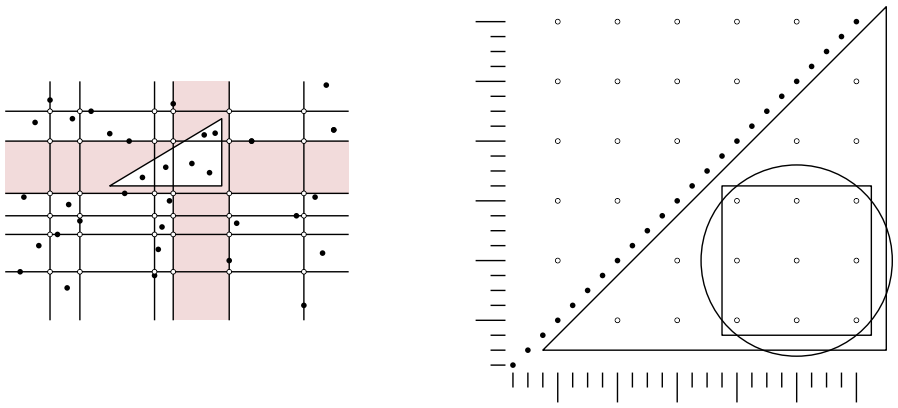
$\text{VC}(\mathcal{V}) \leq 4$: We claim that no 5-element point set $\{a, b, c, d, e\}$ can be shattered by this range space. To see this, consider five points in general position. Let a and d denote the leftmost and rightmost points, respectively. This leaves three remaining points b, c, and e. Let's order these points so that $b_x \leq e_x \leq c_x$. If e lies above the line $\overline{bc}$, then any V-shaped region that contains $\{a, b, c, d\}$ is forced to contain e as well (see Fig. 3(d)). On the other hand, if e lies below the line $\overline{bc}$ then it is impossible to generate the set $\{a, e, d\}$ without including either b or c (see Fig. 3(e)).

Solution 3:

(a) If a rectangle does not contain any point of S, then it lies either in a single column or single row between the grid lines defining S. Therefore, any axis-parallel rectangle that does not hit a point of S contains at most $k-1$ points of P. Right triangles and disks differ only in that they may overlap a single row and a single column (see Fig. 4(a)). Therefore, the number of points of P is at most $2(k-1)$ in each case.

(b) From (a), we know that any axis-parallel rectangle Q that evades S can contain at most $k-1$ points. Such a rectangle has measure $\mu_P(Q) = |Q \cap P|/|P| \leq (k-1)/n$. To make S as small as possible, we should make k as large as possible such that $(k-1)/n < \varepsilon$, or equivalently $k < \varepsilon n + 1$. The largest integer satisfying this is $k \leftarrow \lceil \varepsilon n \rceil$.

A similar analysis for the right triangle and circular disk cases yields $k \leftarrow \lceil \varepsilon n/2 \rceil$.



(a) (b)

Figure 4: Weak ε-nets and samples.

(c) We claim that this construction does not yield ε-samples. Consider the point set shown in Fig. 4(b). For any value of $k \geq 1$, the sampled points fill an $m \times m$ square, where $m = \lfloor n/k \rfloor$. Observe from the figure there exists a rectangle, right triangle, and circular disk that contain at least a quarter of the sample, and yet contain no points of P. Therefore, we have

$$|\mu_S(R) - \mu_P(R)| \;\geq\; \frac{1}{4}.$$

It does not matter how small k is (or how large m is) this approach will not yield an ε-sample for $\varepsilon < 1/4$.

Solution 4:

(a) Let Σ_1 be the range space of all uniform $\pm$-scalings of the right triangle τ. We claim that the VC-dimension is 1.

$\mathrm{VC}(\Sigma_1) \geq 1$: Trivial.

$\mathrm{VC}(\Sigma_1) < 2$: In order to have any chance of being shattered, the points of the set must lie within either the first or third quadrants, and they must all lie within the same quadrant (otherwise no range contains all the points). If the points are sorted by $|x_i + y_i|$, then any range that contains the ith point must contain the entire subset of points before it (see Fig. 5(a)). Thus, no point set of size two or greater can be shattered.

(b)/(c) Let Σ_2 be the range space of all translates, and let Σ_3 be the range space of all homothets of the right triangle. We will show that both have VC-dimension 3. Since $\Sigma_2 \subset \Sigma_3$, it suffices to prove the lower bound for Σ_2 and the upper bound for Σ_3.

$\mathrm{VC}(\Sigma_2) \geq 3$: Consider the point set of Fig. 5(b). The ranges shown cover all the sets of size three and two. Subets of size zero and one are trivial.

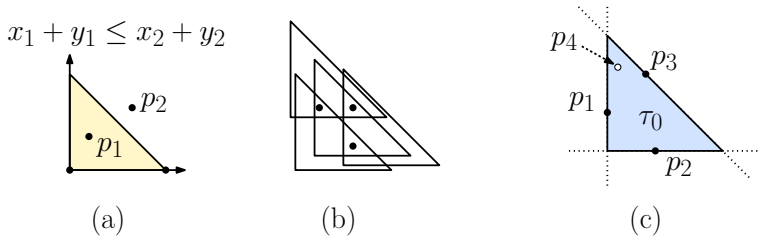


Figure 5: Solution 1: VC-Dimension.

$\mathrm{VC}(\Sigma_3) < 4$: We assert that no four-element point set can be shattered by homothets of the triangle. To see this, consider any set $P = \{p_1, p_2, p_3, p_4\} \in \mathbb{R}^2$. Observe that none of these points can be contained in the convex hull of the others, since otherwise it would be impossible to find a set that separates the point from the others.

Let us assume, therefore, that all four points are on the convex hull. Consider a minimum enclosing triangle τ_0 formed by three lines: a vertical line with all the points of P to its right, a horizontal line with all the points above it, and a diagonal line with all the points below and to the left (see Fig. 5(c)). It is easy to see that τ_0 is a homothet of τ. By general position, least one of the points of P does not lie on τ_0's boundary, say p_4. We assert that *every* homothet of τ that contains $\{p_1, p_2, p_3\}$ must contain τ_0, and therefore, must also contain p_4. Therefore, the set $\{p_1, p_2, p_3\}$ is not in the range space, implying that P is not shattered.

To see the assertion, let p_x, p_y, and p_z denote the points of P that lie on the leftmost, bottommost, and diagonal sides of τ_0, respectively. (Note that it may be that two of these points are equal, since the point may lie on a vertex of τ_0.) Any homothet of τ that contains these points must have its left side to the left of p_x, its bottom side below p_y, and it diagonal side above and right of p_z. Obviously, any such triangle contains τ_0.

Solution to the Challenge Problem: Let Σ_4 be the range space of all translations and rotations of the right triangle. I do not know what the exact VC-dimension of this range space, but the following shows that it is between 5 and 8.

$\mathrm{VC}(\Sigma_4) \geq 5$: Consider a set of five points forming the vertices of the largest regular pentagon that fits within τ such that one of the edges of the pentagon is parallel to the hypotenuse of the right triangle (see Fig. 6(a)). By symmetry, we can exclude cases that are the same up to rotations by multiples of $72°$, the external angle of the pentagon.

This triangle τ generates the set consisting of all five points. The 1-element subsets are easy to generate by translates of τ. Here are the other cases.

- By moving the triangle slightly to the right, we can exclude p_2 to obtain the 4-element subset $\{p_1, p_3, p_4, p_5\}$. By appropriate rotations of $72°$, we can eliminate each of the other four points, thus generating all 4-element subsets.
- By moving the triangle slightly up, we can form the 3-element subset $\{p_1, p_2, p_5\}$, and by moving it down, we can form the 3-element subset $\{p_1, p_3, p_4\}$. All the other 3-element subsets can be generated by appropriate rotations of $72°$.

- By moving this triangle down further still, we can generate the subset $\{p_3, p_4\}$. The other 2-element subsets involving *adjacent pairs* can be generated by by appropriate rotations of 72°. Finally, Fig. 6(b) shows that it is possible to generate the two-element subset $\{p_3, p_5\}$. The other 2-element subsets are again generated by rotations.

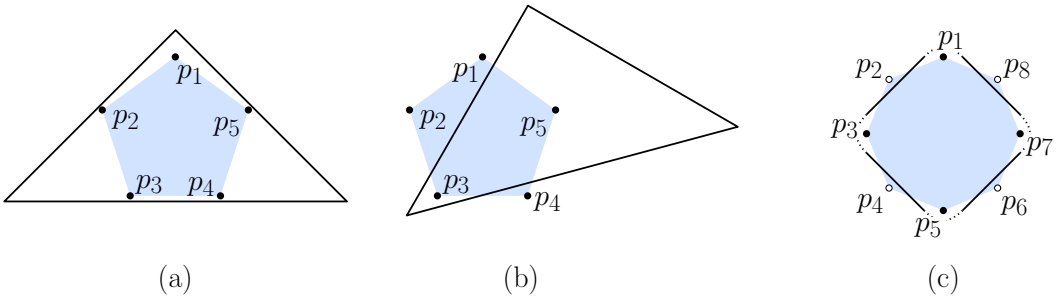


Figure 6: VC-dimension of congruent triangles.

$\mathrm{VC}(\Sigma_4) < 8$: We assert that no 8-element point set can be shattered by rigid motions of the triangle. To see why, we may assume, as before, that the points are in convex position. Number the points in counterclockwise order as $\{p_1, \ldots, p_8\}$. Consider the set that consists of the alternating sequence $\{p_1, p_3, p_5, p_7\}$. The boundary of any convex range that includes these point and excludes the others must cut the edges $\overline{p_i p_{i+1}}$, for $1 \leq i \leq 8$ (with indices wrapping around). See Fig. 6(c). However, it is easy to see that at least four straight line edges are needed to stab these eight edges, which implies that no triangular range can do it. (Note that this upper bound applies to the range space of *all* triangles, so I'm sure that the bound is not tight for the particular problem of rigid motions of a single triangle.)