

## 1. Voronoi diagrams

(a) Proof by induction over the number of points on a circle with radius  $r$  in  $\mathbb{R}^2$ :

**Induction start:**  $n = 3$  Three point lying on a circle with radius  $r$  are non-collinear and the resulting cells are unbounded (figure 1a).

**Induction step:**  $n \rightarrow n + 1$  Let the induction assumption be true for  $n$  points. Let  $p, q$  be two arbitrary neighboring points on the circle. We insert a new point  $o$  exactly at the intersection of the bisector of  $p, q$  and the circle's radius  $r$ . The resulting cells are unbounded (figure 1b).

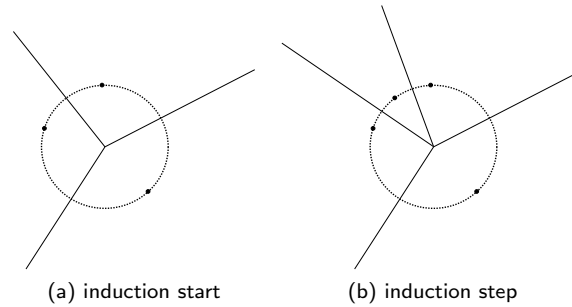


Figure 1: proof by induction over points on a circle

(b) Let  $\text{Vor}(P)$  denote the Voronoi diagram of points  $P$ . Euler's formula for planar graphs states that

$$v - e + f = 2,$$

where  $v, e, f$  denote the number of vertices, edges and faces of the graph.

The number of faces is  $f = |P| = n = \#\text{cells}(\text{Vor}(P))$ . To make use of Euler's formula, we have to construct a planar Graph from  $\text{Vor}(P)$ , by adding an additional vertex to the diagram, situated at infinity, and connecting every infinite edge (from the unbounded cells) to this vertex (figure 2).

We note that each edge in  $\text{Vor}(P)$  has exactly two incident vertices and each vertex of  $\text{Vor}(P) + \infty$  has at least degree (deg) 3. Therefore we conclude

$$\sum_{\text{vertex} \in \text{Vor}(P) + \infty} \deg(\text{vertex}) = 2e \geq 3(v + 1) \quad (1)$$

Since we inserted our additional point  $\infty$ , Euler's Formula for our constructed graph is:

$$(v + 1) - e + f = (v + 1) - e + n = 2$$

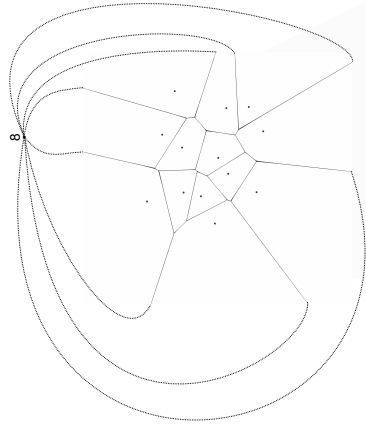


Figure 2: augmenting Voronoi diagram to planar graph

Inserting this into (1) gives us:

$$\begin{aligned} 2e &\geq 3(2 + e - n) = 6 + 3e - 3n \Rightarrow \\ e &\leq 3n - 6 \end{aligned}$$

Inserting this result into (1):

$$v \leq 2n - 5$$

■

## 2. *k*-NN

- (a) We considered the source of the dataset: <https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>. Here it is stated that the categorical always represent some kind of scale and correspond to integer values on that scale. So we change the values given in the input file from letters to integer numbers and the boolean values “true” and “false” to 1 and 0. Now we can apply the euclidian distance function:

$$\delta(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

- (b) See `output_knn.log` and `README.md`.  
(c) See `output_leave_one_out.log` and `README.md`.

■

## 3. Hypothesis Testing

(a)  $n = 100, \#'' + '' = 87$

$$\text{error}_S(h) = \frac{n - \#'' + ''}{n} = \frac{13}{100} = 0.13$$

$$\begin{aligned}\text{error}_{\text{true}}^{95\%}(h) &= \text{error}_S(h) \pm z_N \sqrt{\frac{\text{error}_S(h)(1 - \text{error}_S(h))}{n}} \\ &= 0.13 \pm 1.96 \sqrt{\frac{0.13(1 - 0.13)}{100}} \approx [0.064, 0.196]\end{aligned}$$

(b) With a probability of approximately 95% the true error of the hypothesis  $h$  lies within the above interval.

■