Solutions to Problem Sheet 4

Solution (4.1) For any subset $S \subset \mathcal{X}$ and a set of subsets $\mathcal{A}$, define

$$S_{\mathcal{A}} = \{A \cap S \colon A \in \mathcal{A}\}.$$

(a) Let $S \subset \mathcal{X}$ be a set of cardinality n, with $|S_{\mathcal{C}}| = \Pi_{\mathcal{C}}(n)$, where $\mathcal{C} = \{A \cap B \colon A \in \mathcal{A}, B \in \mathcal{B}\}$. Then the map

$$\begin{aligned} S_{\mathcal{A}} \times S_{\mathcal{B}} &\to S_{\mathcal{C}} \\ (A \cap S, B \cap S) &\mapsto (A \cap S) \cap (B \cap S) = A \cap B \cap S \end{aligned}$$

is, by definition, surjective, and hence

$$\Pi_{\mathcal{C}}(n) = |S_{\mathcal{C}}| \leq |S_{\mathcal{A}} \times S_{\mathcal{B}}| = |S_{\mathcal{A}}| \cdot |S_{\mathcal{B}}| \leq \Pi_{\mathcal{A}}(n) \cdot \Pi_{\mathcal{B}}(n).$$

We used that all the sets involved are finite.

(b) Let $S \subset \mathcal{X}$ be a set of cardinality n, with $|S_{\mathcal{C}}| = \Pi_{\mathcal{C}}(n)$, where $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$. Clearly, $S_{\mathcal{C}} = S_{\mathcal{A}} \cup S_{\mathcal{B}}$, and hence

$$\Pi_{\mathcal{C}}(n) = |S_{\mathcal{C}}| = |S_{\mathcal{A}} \cup S_{\mathcal{B}}| \leq |S_{\mathcal{A}}| + |S_{\mathcal{B}}| \leq \Pi_{\mathcal{A}}(n) + \Pi_{\mathcal{B}}(n).$$

Solution (4.2) (a) Let $d = 1$ and $n \geq 1$. Then there are exactly two ways of labelling the n points on the line, namely one labelling that assigns $+1$ to the positive and -1 to the negative points, and one labelling that assigns -1 to the positive and $+1$ to the negative points. Thus $C(n, 1) = 2$. Also, if $d \geq 1$ is arbitrary and $n = 1$, then there are also exactly two ways of labelling this one points, hence $C(1, d) = 2$ (we used that $\binom{0}{i} = 0$ for $i \geq 1$). Now assume that the formula is true for $n-1$ points in $\mathbb{R}^d$ and for $n-1$ points in $\mathbb{R}^{d-1}$. Suppose we are given a set $S = \{\boldsymbol{x}_i\}_{i=1}^n$ of n points in $\mathbb{R}^d$ such that any subset of at most d points from that set is linearly independent. Any vector $\boldsymbol{v} \in \mathbb{R}^d$ gives rise to a linearly separable labellings by setting $y_i = +1$ if $\boldsymbol{v}^\top \boldsymbol{x}_i > 0$ and $y_j = -1$ if $\boldsymbol{a}^\top \boldsymbol{x}_j < 0$. The hyperplanes orthogonal to $\boldsymbol{x}_i$, $H_i = \{\boldsymbol{x} \colon \boldsymbol{x}^\top \boldsymbol{x}_i = 0\}$, subdivide $\mathbb{R}^d$ into $R(n, d)$ connected components, and two vectors define the same linearly separable labellings if the are in the same region. We thus reduced the problem to that of counting the number of connected components that n hyperplanes subdivide $\mathbb{R}^d$ into, if the normal vectors to these hyperplanes are in "general position" (meaning that no d of them are linearly independent), and $C(n, d) = R(n, d)$. To count the number of regions, note that the number of regions generated by n hyperplanes is the same as the number of regions generated by the first $n-1$ hyperplanes $\{H_1, \dots, H_{n-1}\}$, plus the number of regions added by the n-th hyperplane H_n. This number, in turn, is the same as the number of regions *in* $H_n \cong \mathbb{R}^{d-1}$ that arise by cutting H_n with the $n-1$ hyperplanes $H_1, \dots, H_{n-1}$. Hence, $C(n, d) = C(n-1, d) + C(n-1, d-1)$.

By the induction hypothesis,

$$
\begin{aligned}
C(n,d) = C(n-1,d) + C(n-1,d-1) &= 2\sum_{i=0}^{d-1}\binom{n-2}{i} + 2\sum_{i=0}^{d-2}\binom{n-2}{i} \\
&= 2\sum_{i=0}^{d-1}\left(\binom{n-2}{i} + \binom{n-2}{i-1}\right) \\
&= 2\sum_{i=0}^{d-1}\binom{n-1}{i}.
\end{aligned}
$$

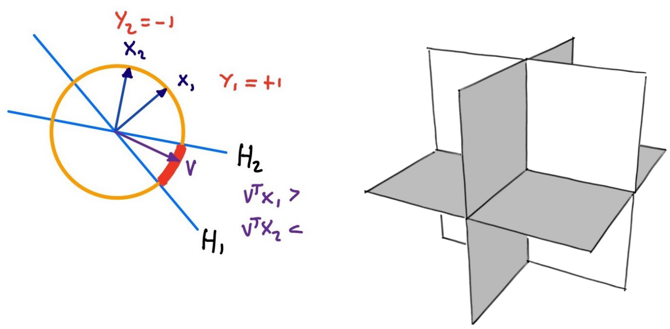


Figure 1: [Left] The correspondence between sign vector configurations and regions of space generated by hyperplanes, in $\mathbb{R}^2$. [Right] Three hyperplanes divide $\mathbb{R}^3$ into $2\left(\binom{2}{0} + \binom{2}{1} + \binom{2}{2}\right) = 8$ regions. If we fix one hyperplane, then this is the number of regions generated by the two other hyperplanes (4), plus the number of these regions that the third hyperplane intersects, as each such intersection generates a new region.

(b) Each function $h \in \mathcal{H}$ corresponds to a linearly separable labelling of the set of points $\{\Phi(x_i)\}_{i=1}^n$. It follows that the number of possible dichotomies is bounded by the number of such labellings in $\mathbb{R}^p$, and the bound follows from (a).

Solution (4.3) Consider a set of $2d+1$ points on a circle and label then with $+1$ and -1. If the number of points labelled with -1 is larger than the number of points labelled with $+1$, then the convex hull of these points is a polygon with at most d sides that contains all the $+1$ points. If the number of $+1$ points is larger, then we take the lines tangent to the points labelled with -1. These intersect in at most d points, and making the resulting shape slightly smaller gives a polygon with at most d sides that contains the $+1$ points.

To show that the VC dimension is *exactly* $2d+1$, one needs to show that *no* configuration of $2d+2$ points can be shattered. First note that if one of the points lies in the convex hull of the other points, then the configuration cannot be shattered, as the point inside the convex hull will always have the same label assigned as the corners of the convex hull. Hence, consider a configuration in which no point is in the convex hull

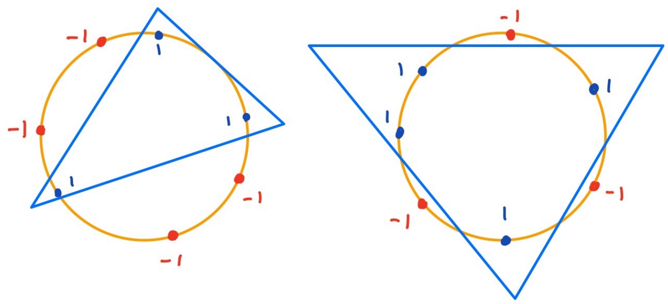


Figure 2: Shattering $2d+1$ points with a d-gon (here, $d=3$). On the left display, the triangle was generated by slightly enlarging the convex hull of the points labelled with $+1$, while on the right display, the labelling was generated by slightly shrinking the triangles generated by the tangent lines to the points labelled with -1.

of the others. This is the same as saying that all $2d+2$ points are the corners (vertices) of a convex $2d+2$-gon. Label the vertices of this polygon by alternating with $+1$ and -1. If we enumerate the vertices in order, starting with a $+1$ labelled vertex, then we get a sequence

$$+-+-+-\cdots+-+-$$

This sequence contains $d+1$ vertices labelled with -1, and separating each of these from its neighbours requires one line segment. Hence, any polygon that separates the $+1$ from the -1 vertices necessarily requires $d+1$ line segments.

Solution (4.4)

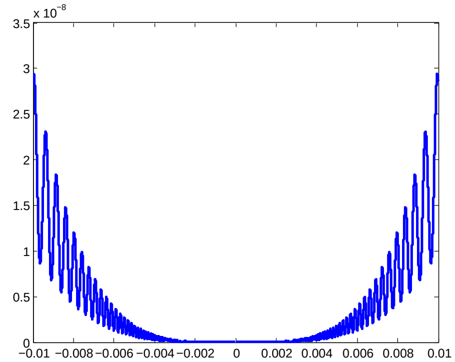
(a) The function $f(x)=x^4$ has a strict minimum at $x=0$, but the second derivative satisfies $f''(0)=0$.

(b) We construct a function that has a strict minimizer x^*, but such that every open neighourhood U of x^* contains other local minimizers. One such function is

$$f(x)=\begin{cases} x^4(\cos(1/x)+2) & \text{for } x\neq 0\\ 0 & \text{for } x=0.\end{cases}$$

We explain the construction of this function:

1. Start out with $g(x)=\cos(1/x)+2$ for $x\neq 0$ and $g(0)=1$. This function has minimizers $x_0=0$ and $x_k=1/(\pi(2k+1)$ for $k\geq 0$, with values $g(x_k)=1$ at all minimizers. Therefore, any open interval around 0 contains (infinitely many) local minimizers x_k other than $x_0=0$.
2. Multipy x^4 to the function: $f(x)=x^4g(x)$. This ensures that $f(0)=0$ and $f(x)>0$ for $x\neq 0$. There are still local minima in every neighbourhood of



0. To see this, compute the derivative

$$f'(x) = x^2(4x\cos(1/x) + \sin(1/x) + 8x). \tag{1}$$

Set $z_m = 1/(\pi/2 + m\pi)$ for $m > 0$. Since $\sin(1/z_m) = \sin(\pi/2 + m\pi) = 1$ for m even and -1 for m odd, and for m sufficiently large the contribution of the other terms is negligible (as the z_m become arbitrary small), the derivative (1) changes signs between successive z_m. Since $f'(x)$ is continuous, it has roots between any z_m and z_{m+1} for large enough m, and these correspond to maxima and minima of f.

The function is in $C^2(\mathbb{R})$. For $x \neq 0$ this is clear, and to verify this at $x = 0$, one shows that the right and left limits as $x \to 0$ of $f'(x)$ and $f''(x)$ coincide (they are in fact 0).

Note the subtle point that one minimizer x^* can have local minimizers that are arbitrary close: while each open interval I surrounding x^* has another local minimizer $\tilde{x}$, every such $\tilde{x}$ has an interval $\tilde{I}$ surrounding it where this $\tilde{x}$ is the only minimizer!

Solution (4.5) The general procedure is as follows: we first make an educated guess as to whether the set could be convex or not. If we think it is not convex, then it is enough to find a *counterexample*: find points in S for which the line segment joining them is not completely contained in S. If we think it is convex, then we can show that for any two points the line segment joining them is in S.

(a) This set is not convex: take $\boldsymbol{x} = (1, 0, 0)^\top$ and $\boldsymbol{y} = (-1, 0, 0)^\top$, then $\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y} = \mathbf{0} \notin S$.

(b) This set is convex: if $\boldsymbol{x}, \boldsymbol{y} \in S$, then $1 \leq x_1 - x_2 < 2$ and $1 \leq y_1 - y_2 < 2$, and

$$\lambda x_1 + (1-\lambda) y_1 - \lambda x_2 - (1-\lambda) y_2 = \lambda(x_1 - x_2) + (1-\lambda)(y_1 - y_2) < \lambda 2 + (1-\lambda) 2 = 2,$$

with the same argument giving the lower bound.

(c) This set is convex. In fact, S is the unit ball of the 1-norm

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|.$$

Given $\boldsymbol{x}, \boldsymbol{y} \in S$,

$$\|\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}\|_1 \leq \lambda \|\boldsymbol{x}\|_1 + (1-\lambda)\|\boldsymbol{y}\|_1 \leq \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1.$$

(d) This set is convex. Here, one needs to show that convex combinations preserve symmetry and positive definiteness of a matrix. The symmetry is clear. As for the positive definiteness, let $\boldsymbol{x} \neq \boldsymbol{0}$ be given. Then

$$\boldsymbol{x}^\top (\lambda \boldsymbol{A} + (1-\lambda)\boldsymbol{B})\boldsymbol{x} = \lambda \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + (1-\lambda)\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x} \geq 0,$$

which shows that positive definiteness is also preserved.