Foundations of data science, summer 2020

JONATHAN LENNARTZ, MICHAEL NÜSKEN, ANNIKA TARNOWSKI

12. Exercise sheet
Hand in solutions until Thursday, 9 July 2020, 12:00

Exercise 12.1 (Exam exercise). (8 points)

Design an exam exercise. $\boxed{8}$

The estimated time to work out a solution should be about 15 minutes. Award credit points to the parts of your exercise reflecting the estimated working time. Make sure that exercise solutions allow to differentiate between a very good student, a good student and a weak student.

Hint: Asking for mere knowledge is not university level.

Exercise 12.2 (Play with the Perceptron algorithm). (10 points)

(i) Pick a vector $w^* \in \mathbb{R}^2$. Pick $1\,000$ random points x_i uniformly distributed in $[-100, 100]^2 \subset \mathbb{R}^2$ with $|\langle w^* \,|\, x \rangle| \geq 1$ and compute labels $\ell_i \leftarrow \text{sign}(\langle w^* \,|\, x_i \rangle) \in \{-1, +1\}$.

  (a) Run the perceptron algorithm (with the kernel function $k(x, y) = x \cdot y$).

  (b) Report about your experiences. $\boxed{3}$

  (c) How does its runtime compare to the lecture's bound on it? $\boxed{2}$

(ii) Pick $1\,000$ random points $\left[\begin{smallmatrix} r\cos\varphi \\ r\sin\varphi \end{smallmatrix}\right]$ with $r \in [0.8, 1.2] \cup [1.6, 2.4]$ uniformly chosen and $\varphi \in [0, 2\pi]$ uniformly chosen and label -1 for $r \leq 1.2$ and label $+1$ for $r \geq 1.6$.

  (a) Run the perceptron algorithm with the Gaussian kernel.

  (b) Report about your experiences. $\boxed{3}$

(iii) Pick $1\,000$ random points $\left[\begin{smallmatrix} 1+r\cos\varphi \\ 1+r\sin\varphi \end{smallmatrix}\right]$ with $r \in [0.8, 1.2] \cup [1.6, 2.4]$ uniformly chosen and $\varphi \in [0, 2\pi]$ uniformly chosen and label -1 for $r \leq 1.2$ and label $+1$ for $r \geq 1.6$.

(a) Run the perceptron algorithm with the Gaussian kernel.

(b) Report about your experiences. 2

Solution (Hints). The Perceptron algorithm with a suitable kernel or after a suitable embedding into $\mathbb{R}^3$ is able to separate the given labels in all the given problems. Only the first part is linearly separable and doable with the basic version. Solutions should provide the necessary material, for example possibly simplified code snippets or images, to understand the reported observations. Submission of full code is only supplementary and cannot replace selection and display of relevant snippets!

The runtime is considerably better than the bound from the lecture. Beware, that the lecture bound has only been proven for the linear perceptron, i.e. with kernel $k(x, y) = \langle x \mid y \rangle$.

The most common mistake in this exercise was choosing a different distribution than given, especially for the first part: A simple way to sample the given distribution is rejection sampling: choose both coordinates uniformly in $[-100, 100]$ and reject the sample if it does not satisfy $|\langle w^* \mid x \rangle| \geq 1$. ○

Exercise 12.3 (VC dimension). (6+3 points)

A hypercuboid in $\mathbb{R}^d$ is a generalized rectangle, ie. subset of the form

$$[a_0, b_0] \times [a_1, b_1] \times \cdots \times [a_{d-1}, b_{d-1}].$$

In $\mathbb{R}^3$ this is called a cuboid.

3 (i) If $\Omega = \mathbb{R}^3$ and $\mathcal{H} = \{\text{cuboid} \subset \mathbb{R}^3\}$, prove that

$$\text{VCdim}(\mathcal{H}) = 6.$$

Solution. At first, we need to show, that there is a set with 6 elements in $\mathbb{R}^3$ that is shattered. Such a set is for example given by

$$\{\pm |e_i\rangle \mid |e_i\rangle \text{ standard unit vector}\}.$$

To see why this shatters, we can at first consider the cube given by $[-0.9, 0.9] \times [-0.9, 0.9] \times [-0.9, 0.9]$. This contains none of the points.

Now when we change one of the coordinates from -0.9 to -1.1 or from 0.9 to 1.1 respectively, we always add a single point to the cube. This is because a point is contained in the cube if and only if each coordinate of the point is contained in the respective interval. So for the given set —as we made sure that 0 is always contained in the intervals we choose— whether a point $\pm |e_i\rangle$ is inside the cube just depends on whether ± 1 is contained in the i-th interval.

Another possibility to show that this shatters is to simply make a case distinction between the ten possible kinds of the 64 subsets —namely up to symmetry— of this set and to show that all can be expressed as a shard.

So we know that $\mathrm{VCdim}(\mathcal{H}) \geq 6$. It remains to show that $\mathrm{VCdim}(\mathcal{H}) \leq 6$, so we need to show that no set A with seven elements shatters.

Let us consider a cube that is large enough such that it contains all given seven points. Now we try to find the smallest cube, that still contains all seven points. So we start to shrink it and move all sides as close to each other as possible. Each side will eventually meet a point that stops it. We call that point a *stopping point*. If there is more than one point stopping one of the sides simultaneously, we choose only one among the possible points. A point can also be designated a stopping point for more than one side.

The cuboid C we have now obtained is the minimal cuboid that contains all points of A. Also, as it can be reconstructed from the stopping points alone, it is the minimal cuboid that contains all the stopping points.

Now consider the set of stopping points S. Notice, that $S \neq A$, as there are at most six stopping points. The claim is that S can never occur as a shard of A, ie. that we can not write S as $A \cap \text{cuboid}$.

This is because any cuboid D that contains S, also contains the minimal cuboid C we defined earlier. But all points of A are already contained in C. So we either obtain a shard that is missing a stopping point or a shard that contains all the points. But as $S \neq A$, we never obtain S as a shard and hence A does not shatter.

Hence no set with seven points shatters, so $\mathrm{VCdim}(\mathcal{H}) = 6$. ○

(ii) Consider $\Omega = \mathbb{R}^d$ and $\mathcal{H} = \left\{\text{hypercuboid} \subset \mathbb{R}^d\right\}$. With methods similar to those seen in class, prove that $\boxed{3}$

$$\mathrm{VCdim}(\mathcal{H}) < 2d + 1.$$

Solution. At first, note that a hypercuboid has $2d$ faces. Visually, there is a upper and lower face for every dimension in the space. Mathematically, a face is given by

$$[a_0, b_0] \times [a_1, b_1] \times \cdots \times \{a_i\} \times \cdots \times [a_{d-1}, b_{d-1}]$$

or

$$[a_0, b_0] \times [a_1, b_1] \times \cdots \times \{b_i\} \times \cdots \times [a_{d-1}, b_{d-1}],$$

where i can be any of the d coordinates.

Now, the solution is exactly the same as the second part in the previous task. You can replace seven with $2d + 1$ and get a set of at most $2d$ stopping points.

You can also take a more mathematical viewpoint when facing these problems. Given a set $A = \{p_0, \dots p_{2d}\}$ the stopping points correspond to

$$\operatorname{argmax}_{x \in A} x_i,$$

where the x_i here denotes the coordinate in direction i, and

$$\operatorname{argmin}_{x \in A} x_i.$$

Again, when there are several points, the argmax and argmin expressions choose one point that maximises the term. In the end we obtain at most $2d$ points as we have $2d$ such expressions.

Call the set of these points S. If we then consider a cuboid C that contains S it is by definition of the form

$$[a_0, b_0] \times [a_1, b_1] \times \cdots \times [a_{d-1}, b_{d-1}].$$

Also we know that $\operatorname{argmax}_{x \in A} x_i \in S$ for all i, so if $S \subset C$ especially $b_i \geq \max\limits_{x \in A} x_i$ for all i. This is again because a point is contained in a cube if and only if each coordinate is contained in the corresponding interval.

Similarly $a_i \leq \min_{x \in A} x_i$. But now all points of A lie in the cube, as all of them satisfy $a_i \leq x_i \leq b_i$.

Since $S \subsetneq A$ is different from A the set S can not be obtained as a shard of A. ○

$\boxed{+3}$ (iii) Can you find a set with $2d$ points in $\mathbb{R}^d$ that shatters? Explain why it does.

Solution. We can again take the set

$$\{\pm |e_i\rangle \mid |e_i\rangle \text{ standard unit vector}\}.$$

Here again we can consider all cuboids of the form

$$[a_0, b_0] \times [a_1, b_1] \times \cdots \times [a_{d-1}, b_{d-1}],$$

with $a_i = -1 \pm 0.1$ and $b_i = 1 \pm 0.1$. For any such cuboid C, we have that $|e_i\rangle \in C$ if and only if $b_i = 1.1$ and $-|e_i\rangle \in C$ if and only if $a_i = -1.1$. So as each of the boundaries of the cube completely determines the state (inside or outside) of exactly one of the given points, we can completely shatter the given set. ◯