Assignment 11

Machine Learning: Algorithms and Theory

Prof. Ulrike von Luxburg / Diego Fioravanti / Moritz Haas
Tobias Frangen / Siavash Haghiri

Summer term 2018 — due to **July 10**

1. When you hand in your assignment you need to hand in the notebook too. Please do not write down a report with just results and/or figures. Ideally you should email it, it is easier to correct and more eco friendly, but we accept printed versions too. From now on not handing in the notebook will result in 0 points for the programming part.
2. Before the end of the course you need to present at least one of your solution in the tutorial. If you do not do that you cannot take the exam! If for any reason you cannot attend let us know, it is possible to change group or find another solution.
3. Join the class on ILIAS otherwise we cannot contact you if we need to.

Exercise 1 (VC dimension I, 1+2+2+4 points) Let $a, b \in \mathbb{R}$, $[a, b]$ denotes the closed interval $\{x \in \mathbb{R} : a \leq x \leq b\}$. Let $\mathbb{1}_{[a,b]}(x)$ be the characteristic function such that:

$$\mathbb{1}_{[a,b]}(x) = \begin{cases} 1 & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

(a) Assume that $\mathcal{F}$ is a finite function class. Prove that

$$\mathrm{VC}(\mathcal{F}) \leq \log_2 |\mathcal{F}|.$$

(b) Consider the function class

$$\mathcal{F} = \{f : \mathbb{R} \to \{0, 1\} \mid f(x) = \mathbb{1}_{[a,b]}(x), a, b \in \mathbb{R}\}.$$

Prove that $\mathcal{N}(\mathcal{F}, n) = 1 + n + \frac{1}{2}n(n-1)$ and $\mathrm{VC}(\mathcal{F}) = 2$.

(c) Consider the function class

$$\mathcal{F} = \{f : \mathbb{R} \to \{0, 1\} \mid f(x) = \sum_{i=1}^{n} \mathbb{1}_{[a_i,b_i]}(x), n \in \mathbb{N}, a_i, b_i \in \mathbb{R}\}.$$

Prove that $\mathrm{VC}(\mathcal{F}) = \infty$.

(d) Consider the function class of hyperplanes, that is

$$\mathcal{F} = \left\{ f : \mathbb{R}^d \to \{0, 1\} \mid f(x) = \frac{1 + \mathrm{sign}(\langle w, x \rangle + b)}{2}, w \in \mathbb{R}^d, b \in \mathbb{R} \right\}.$$

Prove that $\mathrm{VC}(\mathcal{F}) = d + 1$.

Hint: You can use the following theorem (without proving it):

Two sets of points in $\mathbb{R}^d$ can be separated by a hyperplane if and only if the intersection of their convex hulls is empty.

Exercise 2 (VC dimension II, 1+2+2 points) A function class $\mathcal{F}$ of functions defined on a set $\mathcal{X}$ is linearly ordered if it contains at least two functions and if for any $h, g \in \mathcal{F}$ either

$$\forall\, x \in \mathcal{X} : \quad h(x) = 1 \;\Rightarrow\; g(x) = 1$$

or

$$\forall\, x \in \mathcal{X} : \quad g(x) = 1 \;\Rightarrow\; h(x) = 1$$

holds true.

(a) Provide an example of a function class that is linearly ordered.

(b) Prove that $\mathrm{VC}(\mathcal{F}) = 1$ for any function class $\mathcal{F}$ that is linearly ordered.

(c) Suppose that $\mathcal{F}$ is a function class with $\mathrm{VC}(\mathcal{F}) = 1$, containing the functions $g \equiv 0$ and $f \equiv 1$. Prove that $\mathcal{F}$ is linearly ordered.

Exercise 3 (Generalization bound using VC dimension, 1+3+2 points)
In this exercise, you will prove the result providing a generalization bound based on the VC dimension of a (possibly infinite) function class (Theorem 36, Slide 744).

(a) Given a function class $\mathcal{F}$, prove that, for any two independently drawn samples of size n from a probability distribution P and for any function $f \in \mathcal{F}$, the following inequality holds:

$$\forall t > 0, \mathrm{P}\left(R_n(f) - R'_n(f) > t\right) \leq 2e^{-nt^2/2}$$

where $R_n(f)$ and $R'_n(f)$ denote the risks of the function f computed on the two independent samples.

Hint: You can use Hoeffding's inequality (Proposition 28, Slide 703).

(b) Use the result from Part (a) and the symmetrization lemma (Proposition 34, Slide 731) to prove that $\forall\delta$ such that $0 < \delta < 1$, with probability at least $1 - \delta$, all functions $f \in \mathcal{F}$ satisfy:

$$R(f) \leq R_n(f) + 2\sqrt{2\frac{\log(\mathcal{N}(\mathcal{F}, 2n)) + \log(\frac{4}{\delta})}{n}}$$

where $R(f)$ is the true risk of the function f with respect to the distribution P and $\mathcal{N}(\mathcal{F}, 2n)$ is the shattering coefficient.

(c) Finally, use the results from Parts (a) and (b) and Sauer's lemma (Proposition 35, Slide 742) to prove that $\forall\delta$ such that $0 < \delta < 1$, with probability at least $1 - \delta$, all functions $f \in \mathcal{F}$ satisfy:

$$R(f) \leq R_n(f) + 2\sqrt{2\frac{d\log\left(\frac{2en}{d}\right) + \log(\frac{4}{\delta})}{n}}.$$

(Optional) **Exercise 4 (Proving Sauer's Lemma, 3+2 points)**

(a) Sauer's Lemma claims that for any function class $\mathcal{F}$ with $\mathrm{VC}(\mathcal{F}) = d \geq 0$ we have

$$\mathcal{N}(\mathcal{F}, n) \leq \sum_{i=0}^{d} \binom{n}{i}, \quad n \geq 1.$$

Assume that we already proved the lemma for $d = 0$ and arbitrary $n \geq 1$ and also for $n = 1$ and arbitrary $d \geq 0$. Prove the lemma for general $d \geq 0, n \geq 1$ by induction over $d + n$.

(b) Show that

$$\sum_{i=0}^{d} \binom{n}{i} \leq n^d, \quad n \geq d > 1.$$