# Statistical Learning Theory and SVM

Hervé Frezza-Buet

duration : 1h30
Electronic devices are not allowed, all paper documents are allowed

---

**Warning !** All answers have to be justified. Do never answer by only yes or no.

During the academic year, the students of an university have to take $n$ exams. Let $m_i^e \in [0, 100]$ the mark obtained by student $e$ for the exam $i$. In addition to these exams, there is a final test for accessing the next year program. The result of that final test for a student $e$ is denoted by $f^e \in \{\texttt{pass}, \texttt{fail}\}$. We would like to determine from the exams' marks wether the student will pass the final test. We have at our disposal the data set $S = \{(m_1^e, m_2^e, \cdots, m_n^e, f^e)\}_e$ corresponding to the 500 students of the previous year. Let us denote by $|X|$ the number of elements in a set $X$: $|S| = 500$. Let us assume that the students' statistics are the same over the years, and that the exams are the same too.

## Failure prediction from the mean (7 points)

Let us use the average marks of a student in order to know if s/he will pass the test. Let $\mathcal{H}$ be the set of functions $h_\theta$ defined as

$$h_\theta(m_1, m_2, \cdots, m_n) = \begin{cases} \texttt{pass} & \text{if } \frac{1}{n}\sum_i^n m_i \geq \theta \\ \texttt{fail} & \text{otherwise} \end{cases} \tag{1}$$

Let us also use the following notations[1]:

$$\begin{array}{rcl} S^+ & = & \{(m_1, m_2, \cdots, m_n, f) \in S : \ f = \texttt{pass}\} \\ S^- & = & \{(m_1, m_2, \cdots, m_n, f) \in S : \ f = \texttt{fail}\} \\ S_h^{\text{FP}} & = & \{(m_1, m_2, \cdots, m_n, f) \in S : \ h(m_1, m_2, \cdots, m_n) = \texttt{pass} \text{ et } f = \texttt{fail}\} \\ S_h^{\text{FN}} & = & \{(m_1, m_2, \cdots, m_n, f) \in S : \ h(m_1, m_2, \cdots, m_n) = \texttt{fail} \text{ et } f = \texttt{pass}\} \end{array} \tag{2}$$

**Question 1** (1 point) : By using the above set definitions and the $|X|$ notation, write some mathematical expression of the empirical risk of $h_\theta$.

Let us consider the following algorithm, called the $\mathcal{M}$ method, for the setting of a $\theta^\star$ parameter such as $h_{\theta^\star}$ is a good predictor.

- Compute $A = \left\{\frac{1}{n}\sum_i^n m_i\right\}_{(m_1,m_2,\cdots,m_n,f)\in S^+}$ and $B = \left\{\frac{1}{n}\sum_i^n m_i\right\}_{(m_1,m_2,\cdots,m_n,f)\in S^-}$.
- Compute $\mu_A = \frac{1}{|A|}\sum_{a\in A} a$ and $\mu_B = \frac{1}{|B|}\sum_{b\in B} b$.
- $\theta^\star = \frac{\mu_A + \mu_B}{2}$.

**Question 2** (2 point) : Do you think that $\mathcal{M}$ implements a minimization of the empirical risk as an induction principle? If your answer is yes, justify it, otherwise, exhibit a case for which you can find a better $\theta$ (in terms of empirical risk) than the one returned by $\mathcal{M}$.

**Question 3** (2 points) : Propose a method, different from $\mathcal{M}$, that relies on the empirical risk minimization (ERM) on $\mathcal{H}$.

**Question 4** (2 point) : Do you think that there are some methods which compute some $h_\theta$ with good performances while being subject to overfitting ? If you answer yes, give an example, otherwise, justify.

[1] The notation $\{A : \ B\}$ means "the set of $A$s such as $B$".

## Prediction from all the marks, linear SVM (4 points)

Let us use a $\nu$-SVC with the standard dot product in $\mathbb{R}^n$ as a kernel and $\nu = 0.1$. We work with the dataset $S$ defined previously. We measure an 0.01 empirical risk.

**Question 5** (2 points) : Can I affirm that the predictor produced by the SVM will have good performances ? Justify. Is your answer dependent on $n$? on $\nu$?

I apply a cross-validation and find a 0.012 estimated real risk.

I had the final test yesterday and today I got the marks $(m_1, m_2, \cdots, m_n)$ for my exams this year. When fed with my marks, the predictor given by the SVM returns `pass`.

**Question 6** (2 points) : Can I conclude that it is very probable that I will pass the final test ?

## Prediction from all the marks, non-linear SVM (6 points)

Let us assume that, a opposed to the previous question, with the usual dot product for kernel (i.e. the SVM is still linear), the empirical risk is 0.5.

**Question 7** (1 point) : What do you think about the real risk of this SVM ?

Let us continue with a $\nu$-SVC as previously, with $\nu = 0.1$, but the kernel is now a Gaussian kernel with a $\sigma$ parameter.

**Question 8** (1 point) : Using $\sigma = 1$, we compute the empirical risk of the SVM and find 0.01. Is there a chance that the SVM overfit ?

**Question 9** (1 point) : Using $n = 4$ exams, give an approximate value for a suitable $\sigma$.

Let us suppose that the final test only evaluates skills in physics and mathematics. The subjects related to the exams are listed in table 1.

| Mark | Subject |
|---|---|
| $m_1$ | Mathematics |
| $m_2$ | Literature |
| $m_3$ | Sports |
| $m_4$ | Physics |
| $m_5$ | Art |

Table 1: Exams

It is reasonable to assume that marks $m_2, m_3$ et $m_5$ are not related to the student's ability to pass the final test.

**Question 10** (1 point) : Would we get better performances if we restrict the dataset to $\mathbb{R}^2$, i.e. if we work with $S' = \{(m_1, m_4, f)\}_{(m_1,m_2,m_3,m_4,m_5,f)\in S}$ ? Justify.

Let us now suppose that I do not know the subjects related to the marks $(m_1, m_2, m_3, m_4, m_5)$. I consider that some of them may have no relation with the ability to pass the final test.

**Question 11** (2 point) : Describe a procedure for the selection, among the 5 marks, of those which are actually relevant for predicting the final test result.

## In the land of good students (3 points)

Let us now consider that the final test evaluates a skill level that, normally, all the students can reach if they are serious. The last year students were all serious, so *they all pass the final test.* Let us work with the predictors in the set $\mathcal{H}$ defined at the beginning of this text.

**Question 12** (1 point) : Give a value for the $\theta$ parameter such as the real risk, estimated by cross-validation on $s$, is as low as possible.

We would like to detect students who, this year, may not be serious, and thus may fail in the final test.

**Question 13** (2 point) : Can we set up a detector from $S$ ? If the answer is no, justify, otherwise propose a method.