

Mathematical Underpinnings Lab 3 13.03.2024

Task 1 (Vapnik–Chervonenkis dimension)

- a) Let $S_n = \{x_i \in \mathbb{R}^2 : i = 1, 2, \dots, n\}$ be a set of n points in the plane and $\mathcal{F}_{a,b} = \{f : f(x) = a'x + b\}$ be a family of linear functions.
- Draw (using pen and paper) an example of S_2 shattered by $\mathcal{F}_{a,b}$.
 - Draw (using pen and paper) an example of S_3 shattered by $\mathcal{F}_{a,b}$. Is it possible to find an example of three points in \mathbb{R}^2 which are not shattered by $\mathcal{F}_{a,b}$?
 - Is it possible to find an example of S_4 which is not shattered by $\mathcal{F}_{a,b}$?
 - What does it tell you about Vapnik–Chervonenkis dimension of $\mathcal{F}_{a,b}$?
- b) Implement a function which takes d (the dimension of data), n and a set $S_n = \{x_i \in \mathbb{R}^d : i = 1, 2, \dots, n\}$ and returns if the set is shattered by a family $\mathcal{F}_{a,b}$ (if it is linearly separable for any class assignments). To check linear separability you may use:
`sklearn.svm.SVC(C=10000, shrinking=False, kernel='linear', tol=1e-5)`
- c) Estimate the Vapnik–Chervonenkis dimension for the family of linear functions $\mathcal{F}_{a,b}$ for $d = 2, \dots, 10$ by simulation. To do that for each d :
- Take a grid of values of n .
 - For each n sample $N = 50$ sets S_n from multivariate normal distribution.
 - For each sampled set S_n check if the set is shattered (using a function from b)).
 - Draw a conclusion.
- d) Estimate the Vapnik–Chervonenkis dimension for the family of balls in \mathbb{R}^d .
- e) (*) Estimate the Vapnik–Chervonenkis dimension for the classification trees (consider trees without and with pruning).