Aprendizagem 2021

Lab 8: Learning Theory

Practical exercises

I. VC dimension

1. Show graphically what is the VC dimension of the following predictive models:

a) 1-dimensional threshold binary classifiers

$$f(x,a) = \begin{cases} 1 & x \geq a \\ 0 & x < a \end{cases}$$

b) 2-dimensional axis-aligned rectangle binary classifiers

$$f(x_1, x_2;\ a_1, a_2, h, w) = \begin{cases} 1 & a_1 + w \geq x_1 \geq a_1, a_2 + h \geq x_2 \geq a_2 \\ 0 & otherwise \end{cases}$$

c) *2*-dimensional perceptron

$$f(x_1, x_2;\ w_0, w_1, w_2) = sign(w_0 + w_1 x_1 + w_2 x_2)$$

d) *d*-dimensional perceptron

$$f(\mathbf{x};\ \mathbf{w}) = sign(\mathbf{w}^T \mathbf{x})$$

2. Show analytically that the VC dimension of a decision tree on inputs with *d* Boolean features is 2^d.

3. For the following scenarios which would you approximately say has smallest VC dimension?

a) three-dimensional real inputs classified by

i. MLP with one hidden layer with the following units per layer 3 2 2

ii. simple Bayesian classifier with multivariate gaussian likelihood function

b) four-dimensional Boolean inputs classified by

i. decision tree

ii. naive Bayes

c) *N*-dimensional real inputs classified by

i. naive Bayes with Gaussian likelihoods

ii. MLP with two hidden layers with the following units per layer N, $\frac{N}{2}$, $\frac{N}{2}$, 2

iii. simple Bayesian classifier with multivariate gaussian likelihood function

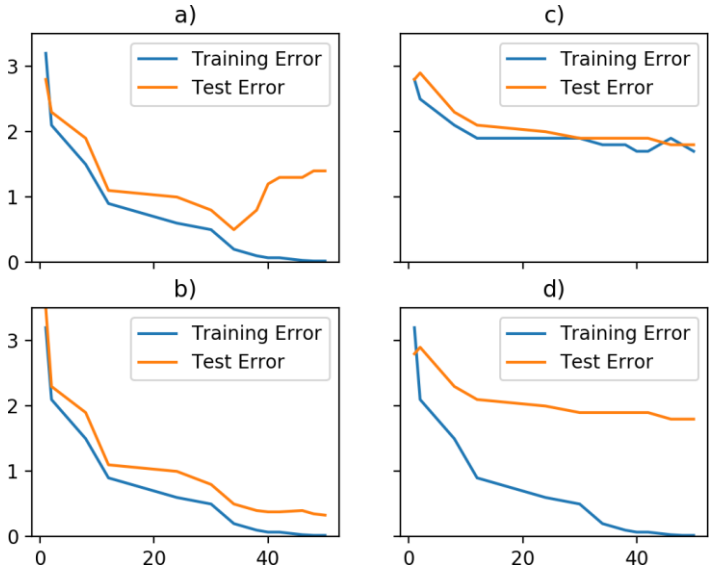
iv. perceptron

II. Model selection

4. Choose between *increase*, *decrease*, *maintain* for each of the following factors:

- training data
- regularization
- VC dimension

For each of the following four scenarios:



Justify each decision.

5. *Thinking question*: Relate the Bias-Variance decomposition and the VC dimension?

Programming quest

6. Using *mlxtend*, compare the bias and variance of a linear regression in the absence and presence of Lasso or Ridge regularization over the housing dataset.
Resource: https://machinelearningmastery.com/calculate-the-bias-variance-trade-off/

7. To assess the number of observations required for a predictive model's ability to generalize:

a) create subsamples of the *housing* dataset with size $n \in \{10,20,..,480,500\}$, 10 times for each n

b) evaluate the training and testing errors of a regression model (e.g. linear regression) at each n

c) are there sufficient observations to guarantee generalization? Discuss