Classification

Jordi Villà i Freixa

Universitat de Vic - Universitat Central de Catalunya
Study Abroad

jordi.villa@uvic.cat

course 2023-2024

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES
UVIC | UVIC-UCC

Índex

Preliminary note

The material in these slides is strongly based on [?]. When other materials are used, they are cited accordingly.

Mathematical notation follows as good as it can a good practices proposal from the Beijing Academy of Artificial Intelligence.

What to expect?

In this session we will discuss:

- Classification methods
- Zero-one loss
- Bayes error rate
- Classification metrics

Regression is a supervised learning method

Supervised methods in which a categorical response variable Y takes one of the possible c values which is to be predicted from a vector of $\mathbf{X}$ explanatory variables, using a prediction function g.

As g classifies the input $\mathbf{X}$ into one of the classes, we call g a classification function or, simply, a *classifier*.

As with any supervised learning technique, the goal is to minimize the expected loss or risk

$$\ell(g) = \mathbb{E}\mathrm{Loss}(Y, g(\mathbf{X})) \tag{1}$$

for some loss function $\mathrm{Loss}(Y, g(\mathbf{X}))$ that quantifies the impact of classifying a response y with $\hat{y} = g(\mathbf{x})$.

Zero-one loss

The zero-noe or *indicator* loss function is the natural choice: $\mathrm{Loss}(y, \hat{y}) := \mathbb{I}\{y \neq \hat{y}\}$: this is: there is no unit loss for a correct classification and a unit loss for wrong one.

This leads to the fact that we aim at taking $g(\mathbf{x})$ to be equal to the class label y for which $\mathbb{P}[Y = y|\mathbf{X} = \mathbf{x}]$ is maximal.

The error we generate in this process is linked to the so-called Bayes error rate.

Pre-classifier

For a given training set τ, a classifier is foten derived from a pre-classifier g_τ, which is a prediction function (learner) that can take any real value, rather than only values in the set of class labels.

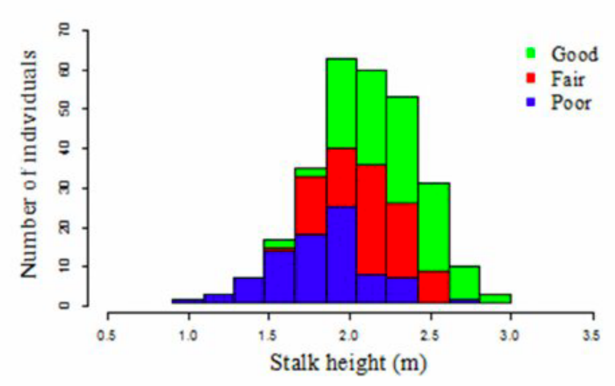

Figure 1: Adapted from here

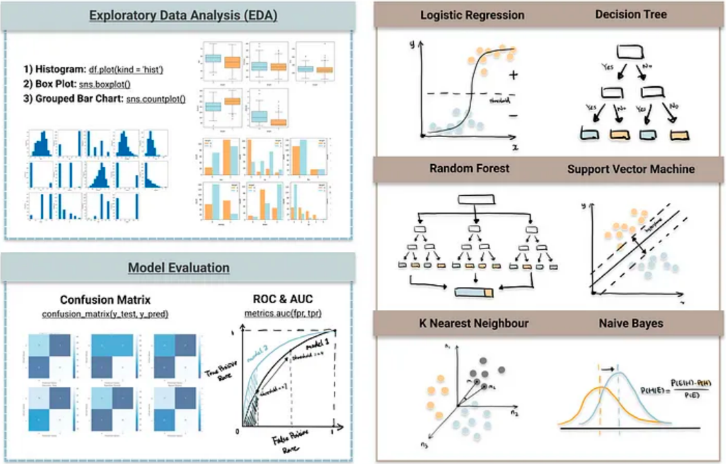

Figure 2: Check the source for a plain explanation of the different classification methods.

Training and test sets. Loss ans confussion matrices

Theoretically, we should be measuring the risk in Eq. **??** and minimizing such equation over some class of functions $\mathscr{G}$. However, as the training loss is often a poor estimate of the risk, this is usually estimated from the test set τ'.

Loss matrix **L** : for the indicator loss function, it contains 0 in the diagonal and 1 everywhere else.

Confusion matrix **M** : counts the number of times that, for the training or test data, the actual (observed) class is i whereas the predicted class is j.

The training/test loss of the classifier in terms of **L** and **M** is $\frac{1}{n}\sum_{i,j}[\mathbf{L} \odot \mathbf{M}]ij$ In the case of the indicator loss, the missclassification error is $1 - \mathrm{tr}(\mathbf{M})/n$

Confusion matrix

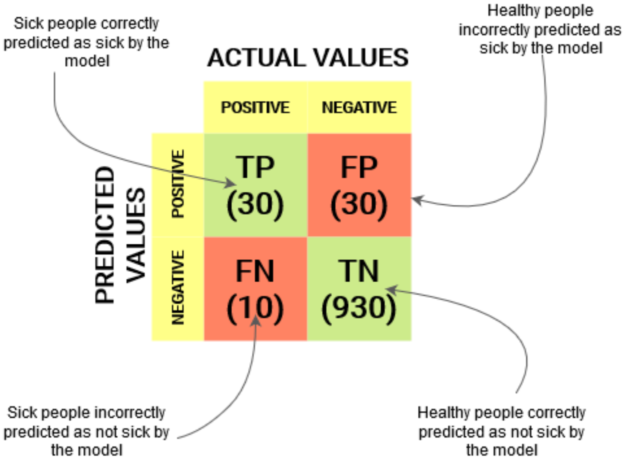

Figure 3: Adapted from here.

Missclassification error and accuracy

In the binary classification case ($c = 2$), and using the indicator loss function, the missclassification error can be written as:

$$\mathrm{error}_j = \frac{\mathrm{fp}_j + \mathrm{fn}_j}{n}$$

and the accuracy can be calculated by measuring the fraction of correctly classified objects:

$$\mathrm{accuracy}_j = 1 - \mathrm{error}_j = \frac{\mathrm{tp}_j + \mathrm{tn}_j}{n}$$

FACULTAT DE CIÈNCIES, TECNOLOGIA I ENGINYERIES
UVIC | UVIC-UCC

We can do better than this in many situations:

- we can modify the loss matrix and make it different from the indicator
- we can modify the the way we measure the classification beyond the accuracy
  - precision: $\mathrm{precision}_j = \frac{\mathrm{tp}_j}{\mathrm{tp}_j + \mathrm{fp}_j}$
  - recall or sensitivity: $\mathrm{recall}_j = \frac{\mathrm{tp}_j}{\mathrm{tp}_j + \mathrm{fn}_j}$
  - specificity: $\mathrm{specificity}_j = \frac{\mathrm{tn}_j}{\mathrm{tn}_j + \mathrm{fp}_j}$
  - F_β score: $F_{\beta,j} = \frac{(\beta^2+1)\mathrm{tp}_j}{(\beta^2+1)\mathrm{tp}_j + \beta^2 \mathrm{fn}_j + \mathrm{fp}_j}$