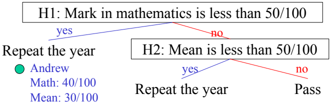
Explaining a Result to the End-User: A Geometric Approach for Classification Problems

I.Alvarez & S.Martin

Geometric information identifies in classification trees the most relevant tests to describe the situation (Alvarez, 04)

Decision tree for next year course applicants



Andrew's trace of classification: mark in math < 50

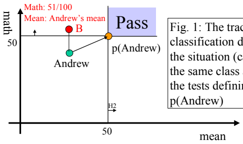


Fig. 1: The trace of Andrew's classification doesn't describe the situation (case B is still in the same class area), contrary to the tests defining case p(Andrew)

Geometric information presents a complementary viewpoint to probabilistic information.

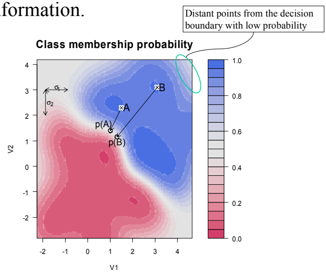
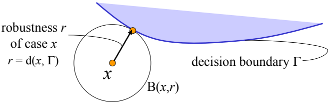
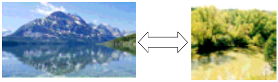


Fig. 2: Points with the same class membership probability can have a very different geometric situation. A is close to the decision boundary, a small change in its attribute values can change the decision.

Geometric approach gives information about the robustness of the decision considering change in the input case



Application to a deterministic classifier for the eutrophication lake problem



oligotrophic lake   eutrophic lake

The state of the lake is defined by the amount of phosphorus and by phosphorus inputs from agriculture (Carpenter et al, 99)

$$\frac{dP(t)}{dt} = -\underbrace{b.P(t)}_{\text{loss}} + L(t) + \underbrace{r.\frac{P^q(t)}{m^q + P^q(t)}}_{\text{recycling from sediments}}$$

Regulation law is a constraint on *dL/dt*.

The viability model gives the set of states where it is possible to maintain the oligotrophic state and agriculture (Martin, 05)

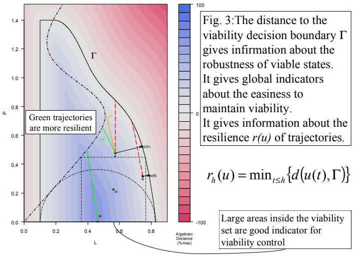


Fig. 3:The distance to the viability decision boundary Γ gives infirmation about the robustness of viable states. It gives global indicators about the easiness to maintain viability. It gives information about the resilience *r(u)* of trajectories.

$$r_h(u) = \min_{t \le h}\{d(u(t), \Gamma)\}$$

Limits: Geometric approach applies to metric space only. Basic metrics:

min – max $\quad y_i = \dfrac{x_i - \min_i}{\max_i - \min_i}$

standard $\quad y_i = \dfrac{x_i - \bar{x}_i}{\sigma_i}$

Computation costs can be exponential (Meijster et al, 00).