Recognition Labs – Minimax Task

Martin Urban, Vojtch Franc, Petr Doubek

October 13, 2009

1 Minimax Task

In one of the previous exercises, we have devised a bayesian desision strategy which minimizes the risk of classification error. Both the conditional probabilities $p(x|y)$ and prior probabilities $p(y)$ were known. In this exercise, we focus on the case when prior probabilities are unknown and bayesian decision strategy is not applicable.

We will show that wrong assumption of prior probabilities can lead to much larger error than accepting the prior probabilites as unknowns and solving the problem as *minimax task*.

1.1 Risk of strategy $d(x)$

Let's assume only two classes $y \in Y = \{1, 2\}$ and 0/1-loss function W. The bayesian risk has form

$$R(d(x)) = \int_X \sum_{y \in Y} p(x)p(y|x)W(d(x), y)dx \quad (1)$$

$$= \int_X \sum_{y:\ y \neq d(x)} p(y)p(x|y)dx \ . \quad (2)$$

Let $X_1 \subset X$ denotes the subset of all $x \in X$ classified by $d(x)$ as class 1 and $X_2 \subset X$ the subset of all x classified as class 2, where $X = X_1 \cup X_2$, $X_1 \cap X_2 = \emptyset$. The equation (2) can be written as sum of two integrals

$$R(d(x)) = \int_{X_1} \sum_{y:\ y \neq d(x)} p(y)p(x|y)dx + \int_{X_2} \sum_{y:\ y \neq d(x)} p(y)p(x|y)dx =$$

$$= p(y = 2) \int_{X_1} p(x|y = 2)dx + p(y = 1) \int_{X_2} p(x|y = 1)dx \ .$$

.

By substituting $p(y = 2) = 1 - p(y = 1)$ we gain a risk function

$$R(d(x)) = p(y = 1) \left[\int_{X_2} p(x|y = 1)dx - \int_{X_1} p(x|y = 2)dx \right] + \int_{X1} p(x|y = 2)dx \quad (3)$$

which is linear for fixed strategy $d(x)$ with respect to prior probability $p(y = 1)$ and thus to $p(y = 2)$ as well. Remember that the conditional probabilities

$p(x|y=i)$ are given and for fixed $d(x)$ (and thus fixed sets X_1, X_2) the integrals are constant.

Moreover, from the risk function (3) also follows that if we find for given $p(x|y=1)$ a $p(x|y=2)$ such strategy $d(x)$ that

$$\int_{X_2} p(x|y=1)dx - \int_{X_1} p(x|y=2)dx = 0 \ , \tag{4}$$

then the risk is independent of prior probabilities $p(y=i)$ and equal to

$$R(d(x)) = \int_{X_1} p(x|y=2)dx = \int_{X_2} p(x|y=1)dx \ .$$

1.2 Minimax decision

The task is to find a decision strategy $d(x)$ given the conditional probabilities $p(x|y=1)$ and $p(x|y=2)$ but with prior probabilities $p(y=1)$ and $p(y=2)$ unknown. We will seek such decision strategy $d(x)$ whose worst case error for any $p(y=1)$ is minimal, i.e. minimizing the maximal error

$$d_m(x) = \operatorname*{argmin}_{d(x)} \max_{p(y=1)} R\left(d(x), p(y=1)\right) \ . \tag{5}$$

1.2.1 Properties of minimax decision

- Minimization of worst case error

  $$\operatorname*{argmin}_{d(x)} \max_{p(y=1)} R\left(d(x), p(y=1)\right) \tag{6}$$

  is equivalent to minimization

  $$\operatorname*{argmin}_{d(x)} \max \left\{ \int_{X_2} p(x|y=1)dx, \int_{X_1} p(x|y=2)dx \right\} \ . \tag{7}$$

- The risk of minimax decision $d_m(x)$ cannot be lower than the risk of worst bayesian decision strategy, i.e. bayesian strategy for the least favourable prior probability $p(y=1)$.

- If for given $p(x|y=1)$, $p(x|y=2)$ exists bayesian decision strategy $d_b(x)$ such that

  $$\int_{X2} p(x|y=1)dx = \int_{X_1} p(x|y=2)dx \ , \tag{8}$$

  then $d_b(x)$ is also a minimax solution and it is a bayesian strategy with the maximal risk w.r.t. $p(y=1)$, i.e. bayesian strategy for the least favourable prior probabilities $p(y=1)$, $p(y=2)$.

Proofs omitted.