Non-Bayesian Decision Making

Boris Flach

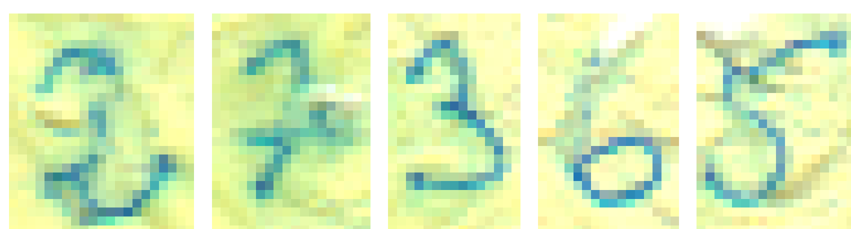
Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
http://cmp.felk.cvut.cz/~flachbor, flachbor@cmp.felk.cvut.cz

1. Bayesian decision making (reconsidered)

Check yourself: have you understood the principles of Bayesian decision making? Answer the following two questions.

- An individual has been described by a neighbour as follows: "Steve is very shy and withdrawn, invariably helpful but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." Is Steve more likely to be a librarian or a farmer?

- You are given ℓ images $x_1, x_2, \ldots, x_\ell$ of digits. You should decide on their sum s. The loss function is $W(s, s') = (s - s')^2$. An OCR algorithm is available for this purpose. It returns the posterior probabilities $p_{K|X}(k \mid x_i)$, $k = 0, \ldots, 9$ for each of the images. What is the optimal decision on s?



2. When do we need non-Bayesian decisions?

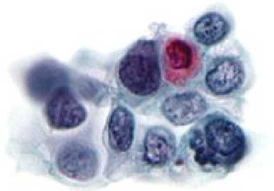
Ingredients & pre-requisites of Bayesian decision making:

- feature space X, (hidden) state space K, decision space D
- real valued loss function $W\colon K \times D \to \mathbb{R}$
- $x \in X$ and $k \in K$ are random events, joint probability

$$p_{XK}(x,k) = p_{X|K}(x \mid k)\, p_K(k) = p_{K|X}(k \mid x)\, p_X(x)$$

A. Can you define a reasonable **loss function** in the following cases?

- automated ZIP-code recognition (OCR). K-set of all ZIP-codes, $D = K \cup \{\text{reject}\}$. "reject" means "a human shall decide..."
- automated cervical cancer screening, $K = \{\text{pre-cancer, healthy}\}$, $D = \{\text{NAD, check up nec.}\}$



- nuclear reactor, K={safe mode, dangerous state}, $D = K$

2. When do we need non-Bayesian decisions?

B. Are the hidden states $k \in K$ **random events** (i.e. can we assign probabilities $p_K(k)$?) in the following cases?

- automated ZIP-code recognition (OCR). K-set of all ZIP-codes
- isolated word speech recognition for a service robot, K-vocabulary
- nuclear reactor, K={safe mode, dangerous state}

C. Are the observed features $x \in X$ given the hidden state $k \in K$ random events? Can we assign probabilities $p_{X|K}(x \mid k)$ to them? Consider the following case

- The service robot (mentioned above) is controlled by a fixed set of speakers $s \in S$. If $x \in X$ denotes the audio signal and $k \in K$ denotes the word (class), then $p_{X|SK}(x \mid s, k)$ is a (conditional) probability. But $s \in S$ is not necessarily random!

Conclusion: We need different decision strategies if the criteria for Bayesian decision making are not met!

3. Formulation of non-Bayesian tasks

A. Neyman-Pearson task

- observations $x \in X$, hidden states $k = 1$ normal, $k = 2$ dangerous, i.e., $K = \{1, 2\}$.
- p.d.s $p_{X|K}(x \mid k)$ are known
- decision strategy: given x decide if the object is in normal or dangerous state, i.e.
  - partition X into two subsets $X_1 \cap X_2 = \emptyset$, $X_1 \cup X_2 = X$ or, more general,
  - $\alpha_{1,2} \colon X \to [0, 1]$, where $\alpha_1(x) + \alpha_2(x) = 1$, $\forall x \in X$
- each strategy is characterised by two numbers

$$\sum_{x \in X_2} p_{X|K}(x \mid 1) = \sum_{x \in X} \alpha_2(x) p_{X|K}(x \mid 1) \qquad \text{false alarm}$$

$$\sum_{x \in X_1} p_{X|K}(x \mid 2) = \sum_{x \in X} \alpha_1(x) p_{X|K}(x \mid 2) \qquad \text{overlooked danger}$$

Task: Choose the strategy which minimises the probability of false alarm subject to: the probability of overlooked danger is less then ϵ.

3. Formulation of non-Bayesian tasks

Neyman, Pearson (1928,1933) optimal strategy decides based on

$$\frac{p_{X|K}(x \mid 1)}{p_{X|K}(x \mid 2)} \lesseqgtr \theta$$

where θ is some threshold.

3. Formulation of non-Bayesian tasks

B. MiniMax task

- as in Neyman-Pearson task – no loss function, hidden states are non-random
- p.d.s $p_{X|K}(x \mid k)$ are known
- in contrast to N-P, hidden states $k \in K$ are symmetric

The decision strategy

- partitions X into $|K|$ subsets, $\cup_{k \in K} X_k = X$, $X_k \cap X_{k'} = \emptyset$ or, more general,
- $\alpha_k \colon X \to [0, 1]$, where $\sum_{k \in K} \alpha_k(x) = 1$, $\forall x \in X$

and is characterised by $|K|$ numbers (error probabilities)

$$\omega_k(\alpha) = \sum_{x \notin X_k} p_{X|K}(x \mid k) = \sum_{x \in X} \big(1 - \alpha_k(x)\big)\, p_{X|K}(x \mid k)$$

Task: Choose the strategy which minimises the maximum of these numbers

$$\alpha^* = \arg\min_{\alpha \in \mathcal{A}} \max_{k \in K} \omega_k(\alpha)$$

3. Formulation of non-Bayesian tasks

The optimal decision strategy is

$$k^* = \operatorname*{arg\,max}_{k \in K} \big[\tau_k \, p_{X|K}(x \mid k)\big]$$

where τ_k, $k \in K$ are some non-negative weights.

C. Wald task

Generalise the Minimax task by allowing for rejection (i.e. introduce X_0 or α_0). Each strategy is now characterised by (modified) numbers ω_k and numbers

$$\chi_k = \sum_{x \in X_0} p_{X|K}(x \mid k) = \sum_{x \in X} \alpha_0(x) \, p_{X|K}(x \mid k)$$

Task: minimise the highest rejection probability, i.e., $\max_{k \in K} \chi_k$
subject to: all misclassification probabilities are less then some ϵ, i.e., $\omega_k < \epsilon$, $\forall k \in K$.

The optimal decision strategy for the case $|K| = 2$: compare the likelihood ratio

$$\gamma(x) = \frac{p_{X|K}(x \mid 1)}{p_{X|K}(x \mid 2)}$$

with two thresholds θ_1 and θ_2.

4. Solving non-Bayesian tasks

How to find the optimal decision strategy for a particular task of non-Bayesian decision making?

All these tasks are usually **linear optimisation tasks**. Apply duality for LP and consider **complementary slackness**.

Example: Neyman-Pearson task

$$
\begin{aligned}
\text{minimise} \quad & \sum_{x \in X} p_{X|K}(x \mid 1)\, \alpha_2(x) & \\
\text{subject to} \quad & -\sum_{x \in X} p_{X|K}(x \mid 2)\, \alpha_1(x) \geqslant -\epsilon & \mid \tau \geqslant 0 \\
& \alpha_1(x) + \alpha_2(x) = 1, \quad \forall x \in X & \mid t(x) \\
& \alpha_1(x), \alpha_2(x) \geqslant 0, \quad \forall x \in X &
\end{aligned}
$$

4. Solving non-Bayesian tasks

The dual task reads

$$
\begin{aligned}
\text{maximise} \quad & \sum_{x \in X} t(x) - \epsilon \tau \\
\text{subject to} \quad & t(x) - p_{X|K}(x \mid 2)\, \tau \leqslant 0, \quad \forall x \in X && \mid \alpha_1(x) \geqslant 0 \\
& t(x) \leqslant p_{X|K}(x \mid 1), \quad \forall x \in X && \mid \alpha_2(x) \geqslant 0 \\
& \tau \geqslant 0
\end{aligned}
$$

and complementary slackness

$$
\begin{aligned}
\alpha_1^*(x) \big[\tau^* \, p_{X|K}(x \mid 2) - t^*(x) \big] &= 0 \quad \forall x \in X \\
\alpha_2^*(x) \big[p_{X|K}(x \mid 1) - t^*(x) \big] &= 0 \quad \forall x \in X
\end{aligned}
$$

where the asterisk is used to denote the solution of the primal and dual task.

4. Solving non-Bayesian tasks

It follows:

$$t^*(x) = \min\left[p_{X|K}(x \mid 1),\; \tau^*\, p_{X|K}(x \mid 2)\right]$$

and the **optimal decision** reads

- decide for $k = 1$ (i.e. $\alpha_1^*(x) = 1$)) if $p_{X|K}(x \mid 1) > \tau^*\, p_{X|K}(x \mid 2)$
- decide for $k = 2$ (i.e. $\alpha_2^*(x) = 1$)) if $p_{X|K}(x \mid 1) < \tau^*\, p_{X|K}(x \mid 2)$

i.e. the decision is made based on the liklehood ratio

$$\gamma(x) = \frac{p_{X|K}(x \mid 1)}{p_{X|K}(x \mid 2)} \lessgtr \tau^*$$

More details in Chapter 2 of

Schlesinger M.I., Hlaváč V.: Ten lectures on statistical and structural pattern recognition. Kluwer Academic Publisher, Dordrecht, The Netherlands, 2002, 519 p.