Classifiers 2

Z. Straka, P. Švarný, J. Kostlivá

Today two examples:

1. Covid-19 testing example
2. Strange loss function for classification

Covid-19 testing example

Covid-19 testing example

Source: Seznam zpravy
For details see a post from Jakub Steiner on Facebook.

Covid-19 testing example

Let's suppose that 0.5% of a population has already been infected by covid-19. Someone else bought covid-19 tests with specificity=0.9 (specificity $= \frac{TN}{TN+FP}$) and wants to test 2000 people from the population. How many of the tests will be false positive?

A: 10

B: 99

C: 199

D: 399

Covid-19 testing example

Let's suppose that 0.5% of a population has already been infected by covid-19. Someone else bought covid-19 tests with specificity=0.9 (specificity $= \frac{TN}{TN+FP}$) and wants to test 2000 people from the population. How many of the tests will be false positive?

C: $199 = (0.995 * 2000) * 0.1$

- $0.995 * 2000 = 1990$ people have not been infected
- As specificity is 0.9, ten percent of all negative samples (i.e., TN+FP) are determined as false positive.

Strange loss function for classification

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

Let's solve it together. Step by step :)

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

How to start?

A: $r(\delta) = \sum_x \sum_s l(s, x) P(x, s)$
B: $r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s)$
C: $r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(\delta(x), s)$
D: $r(\delta) = \sum_s l(s, \delta(x)) P(\delta(x), \delta(s))$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s,d) = K, \quad d = s,$
$l(s,d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

How to start?

B: $r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s) = ?$$

A: $\sum_x P(x) \sum_s l(s, \delta(x)) P(s|x)$
B: $P(x) \sum_s l(s, \delta(x)) P(s|x)$
C: $\sum_x P(x) \sum_s l(s, \delta(x)) P(x|s)$
D: $\sum_x P(x) \sum_s l(s, \delta(x)) P(s, x)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s) =?$$

A: $\sum_x P(x) \sum_s l(s, \delta(x)) P(s|x)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s) = \sum_x P(x) \sum_s l(s, \delta(x)) P(s|x)$$

Optimal strategy for given x:

- A: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(x|s)$
- B: $\delta^*(x) = \arg\max_d \sum_s l(s, d) P(s|x)$
- C: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$
- D: $\delta^*(x) = \arg\max_d \sum_s l(s, d) P(d|x)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$r(\delta) = \sum_x \sum_s l(s, \delta(x)) P(x, s) = \sum_x P(x) \sum_s l(s, \delta(x)) P(s|x)$$

Optimal strategy for given x:

C: $\delta^*(x) = \arg\min_d \sum_s l(s, d) P(s|x)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \text{arg min}_d \sum_s l(s, d)P(s|x) = ?$$

A: arg $\min_d \left(P(s = d|x)K + \sum_{s \neq d} P(s|x)\right)$
B: arg $\max_d \left(P(s = d|x) + \sum_{s \neq d} P(s|x)K\right)$
C: arg $\min_d \left(P(s = d|x) + \sum_{s \neq d} P(s|x)K\right)$
D: arg $\min_d \left(P(s \neq d|x)K + \sum_{s \neq d} P(s|x)\right)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:

$l(s, d) = K, \quad d = s,$

$l(s, d) = 1, \quad d \neq s.$

What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \arg\min_d \sum_s l(s, d)P(s|x) = ?$$

A: $\arg\min_d \left(P(s = d|x)K + \sum_{s \neq d} P(s|x)\right)$

See:

$$\sum_s l(s, d)P(s|x) = l(s_1, d)P(s_1|x) + l(s_2, d)P(s_2|x) + \dots$$

$$+l(s_k = d, d)P(s_k = d|x) + \dots + l(s_n, d)P(s_n|x) = 1P(s_1|x) + 1P(s_2|x) + \dots + KP(s_k = d|x) +$$

$$\dots + 1P(s_n|x) = P(s = d|x)K + \sum_{s \neq d} P(s|x)$$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s,d) = K, \quad d = s,$
$l(s,d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \text{arg min}_d \sum_s l(s,d)P(s|x) = \text{arg min}_d \left(P(s=d|x)K + \sum_{s\neq d} P(s|x)\right) = ?$$

A: $\text{arg min}_d \left(P(s=d|x)K + (P(s=d|x) - 1)\right)$
B: $\text{arg max}_d \left(P(s=d|x) + \sum_{s\neq d} P(s|x)K\right)$
C: $\text{arg max}_d \left(P(s=d|x)K + (1 - P(s=d|x))\right)$
D: $\text{arg min}_d \left(P(s=d|x)K + (1 - P(s=d|x))\right)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:

$l(s,d) = K, \quad d = s,$

$l(s,d) = 1, \quad d \neq s.$

What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \text{arg min}_d \; (\sum_s l(s,d)P(s|x)) = \text{arg min}_d \; (P(s=d|x)K + \sum_{s \neq d} P(s|x)) = ?$$

D: $\text{arg min}_d \; (P(s=d|x)K + (1 - P(s=d|x)))$

Notice that:

$$\sum_{s \neq d} P(s|x) + P(s=d|x) = 1$$

$$\sum_{s \neq d} P(s|x) = 1 - P(s=d|x)$$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:

$l(s,d) = K, \quad d = s,$

$l(s,d) = 1, \quad d \neq s.$

What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \arg\ \min_d \sum_s l(s,d)P(s|x) = \arg\ \min_d P(s=d|x)K + \sum_{s\neq d} P(s|x) =$$

$$= \arg\ \min_d \left(P(s=d|x)K + (1 - P(s=d|x))\right) = ?$$

A: $\arg\ \max_d \left(P(s=d|x)K + (1 - P(s=d|x))\right)$

B: $\arg\ \min_d \left(P(s=d|x)(K-1)\right)$

C: $\arg\ \max_d \left(P(s=d|x)K\right)$

D: $\arg\ \min_d \left(P(s=d|x)(K+1)\right)$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s,d) = K, \quad d = s,$
$l(s,d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \text{arg min}_d \sum_s l(s,d)P(s|x) = \text{arg min}_d \left(P(s=d|x)K + \sum_{s \neq d} P(s|x)\right) =$$

$$= \text{arg min}_d \left(P(s=d|x)K + (1 - P(s=d|x))\right) =?$$

B: $\text{arg min}_d \left(P(s=d|x)(K-1)\right)$

See:

$$\text{arg min}_d P(s=d|x)K + (1 - P(s=d|x)) = \text{arg min}_d P(s=d|x)(K-1) + 1 =$$

$$= \text{arg min}_d P(s=d|x)(K-1)$$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\cdots = \text{arg min}_d \; (P(s = d|x)K + (1 - P(s = d|x))) = \text{arg min}_d \; (P(s = d|x)(K - 1))$$

What values must be K to $\text{arg min}_d \; (P(s = d|x)(K - 1)) = \text{arg max}_d \; P(s = d|x)$? Select the most general option.

A: $K \leq 0$

B: $K < 1$

C: $K < 2$

D: $K \leq 2$

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s, d) = K, \quad d = s,$
$l(s, d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\cdots = \text{arg min}_d \ (P(s = d|x)K + (1 - P(s = d|x))) = \text{arg min}_d \ (P(s = d|x)(K - 1))$$

What values must be K to arg $\min_d \ (P(s = d|x)(K-1)) = \arg\max_d \ P(s = d|x)$? Select the most general option.

B: $K < 1$

For $K < 1$ value of $(K - 1)$ is negative. Therefore, minimization $\min_d \ (P(s = d|x)(K - 1))$ is equivalent to maximization $\max_d \ P(s = d|x)$.

Strange loss function for classification

Consider a Bayesian decision task in which the set of states is the same as the set of decisions (i.e., S=D) and the loss function is defined as:
$l(s,d) = K, \quad d = s,$
$l(s,d) = 1, \quad d \neq s.$
What values must be K to use $\delta^*(x) = \arg\max_d p(d|x)$ to find the optimal decision?

$$\delta^*(x) = \text{arg min}_d \sum_s l(s,d)P(s|x) = \text{arg min}_d \left(P(s=d|x)K + \sum_{s \neq d} P(s|x)\right) =$$

$$= \text{arg min}_d \left(P(s=d|x)K + (1 - P(s=d|x))\right) = \text{arg min}_d \left(P(s=d|x)(K-1)\right) =$$

For $K < 1$:

$$= \text{arg max}_d\ P(s=d|x) = \text{arg max}_d\ P(d|x)$$

:)