

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

The image shows a screenshot of a news article from the website Seznam Zpravy. At the top left, there is a logo 'SZ' and the text 'Seznam Zpravy'. To the right is a search bar with the placeholder text 'Hledat...'. Below the search bar, the main header features a red and white coronavirus particle icon, the word 'Koronavirus', and a red box containing 'R 0,9' with a question mark icon. To the right of this, it says 'ČR: Testů 348,849' and 'Nakaže'. Below the header is a navigation menu with links: 'ONLINE', 'ČESKO', 'SVĚT', 'MAPA', 'ZÁCHRANA BYZNYSU', and 'NEWSLETTER'. The main content area has a breadcrumb trail: 'Zprávy » Koronavirus » Testy » Přesné testy odhalily, že v Česku už měl koronavirus každý dvacátý'. The main headline is 'Přesné testy odhalily, že v Česku už měl koronavirus každý dvacátý'.

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?

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

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?

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

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

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?

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

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

B: $\arg \max_d (P(s = d|x) + \sum_{s \neq d} P(s|x)K)$

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

D: $\arg \min_d (P(s \neq d|x)K + \sum_{s \neq d} P(s|x))$

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?

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

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

See:

$$\begin{aligned} \sum_s l(s, d)P(s|x) &= l(s_1, d)P(s_1|x) + l(s_2, d)P(s_2|x) + \dots \\ &+ l(s_k = d, d)P(s_k = d|x) + \dots + l(s_n, d)P(s_n|x) = 1P(s_1|x) + 1P(s_2|x) + \dots + KP(s_k = d|x) + \\ &\dots + 1P(s_n|x) = P(s = d|x)K + \sum_{s \neq d} P(s|x) \end{aligned}$$

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?

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

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

B: $\arg \max_d (P(s = d|x) + \sum_{s \neq d} P(s|x)K)$

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

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

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?

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

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

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?

$$\begin{aligned} \delta^*(x) &= \arg \min_d \sum_s l(s, d)P(s|x) = \arg \min_d P(s = d|x)K + \sum_{s \neq d} P(s|x) = \\ &= \arg \min_d (P(s = d|x)K + (1 - P(s = d|x))) =? \end{aligned}$$

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

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

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

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

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?

$$\begin{aligned} \delta^*(x) &= \arg \min_d \sum_s l(s, d)P(s|x) = \arg \min_d (P(s = d|x)K + \sum_{s \neq d} P(s|x)) = \\ &= \arg \min_d (P(s = d|x)K + (1 - P(s = d|x))) =? \end{aligned}$$

$$\text{B: } \arg \min_d (P(s = d|x)(K - 1))$$

See:

$$\begin{aligned} \arg \min_d P(s = d|x)K + (1 - P(s = d|x)) &= \arg \min_d P(s = d|x)(K - 1) + 1 = \\ &= \arg \min_d P(s = d|x)(K - 1) \end{aligned}$$

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?

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

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.

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?

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

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?

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

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

For $K < 1$:

$$= \arg \max_d P(s = d|x) = \arg \max_d P(d|x)$$

:)