

Name:

1. (6 points) Bayes classification.

Assume there are two classes of mushrooms, edible (E) and toxic (T). An automatic system designed to help with recognizing the two classes outputs a discrete feature which is 0 (very low poisoning risk), 1 (moderate poisoning risk) or 2 (high poisoning risk).

The prior probabilities for the two classes are

$$p(E) = 0.99, p(T) = 0.01.$$

The conditional probabilities of system output given the true mushroom classes are

$$\begin{aligned} p(0|E) &= 0.89, & p(1|E) &= 0.10, & p(2|E) &= 0.01 \\ p(0|T) &= 0.01, & p(1|T) &= 0.20, & p(2|T) &= 0.79. \end{aligned}$$

- (a) (2p) Compute the conditional probability $p(T|1)$ of the mushroom being toxic given the system reports moderate risk.
- (b) (2p) The goal is to make a binary decision Y (yes, mushroom is accepted) or N (no, mushroom is thrown away). Consider the penalty function

$$\begin{aligned} W(E, Y) &= 0, & W(E, N) &= 1, \\ W(T, Y) &= 1000, & W(T, N) &= 0. \end{aligned}$$

Compute the Bayes-optimal decision strategy, ie. the optimal decision for automatic system outputs 0, 1 and 2.

- (c) (2p) In addition to what you just did, give a short formal summary (using formulas) of how the optimal Bayesian strategy is constructed.
- 2. (6 points) Parameter estimation.

A dice is rolled repeatedly a number of times until a ‘six’ appears on the upper side. The probability of that happening at the n -th dice throw is

$$p(n) = (1 - q)^{n-1}q, \quad (1)$$

where q is the parameter representing the probability of ‘six’ happening.

- (a) (1p) Justify the form of $p(n)$ (explain why the formula is correct.)
- (b) (3p) What is the maximum likelihood estimate of q for a dice that showed “six” at throw k .
- (c) (2p) Assume we set the limit on the number of times the dice is rolled. Let this maximum number of dice throws be n_{max} . What is the maximum likelihood estimate for q when n_{max} is reached without ‘six’ showing up?
- 3. (4 points) Neural Networks.

- (a) (1p) Explain what is the vanishing gradient problem and how to avoid it.
- (b) (1p) What is max-pooling? Why is it used?
- (c) (1p) What is softmax?
- (d) (1p) What is stochastic gradient descent?
- 4. (4 points) Explain the algorithm of K-means++. What is the reason of using K-means++ over standard K-means?

Jméno:

1. (6b) Bayesovská klasifikace.

Předpokládejme, že máme dvě třídy hub, jedlé (E) a toxické (T). Automatický systém navržený tak, aby pomohl s rozpoznáním těchto tříd, poskytuje výstup, který je 0 (velmi nízké riziko otravy), 1 (střední riziko otravy) nebo 2 (vysoké riziko otravy).

Apriorní pravděpodobnosti pro uvedené dvě třídy jsou

$$p(E) = 0.99, p(T) = 0.01.$$

Podmíněné pravděpodobnosti jsou

$$\begin{aligned} p(0|E) &= 0.89, & p(1|E) &= 0.10, & p(2|E) &= 0.01 \\ p(0|T) &= 0.01, & p(1|T) &= 0.20, & p(2|T) &= 0.79. \end{aligned}$$

- (a) (2b) Vypočítejte podmíněnou pravděpodobnost $p(T|1)$, že houba je toxická za předpokladu, že systém hlásí střední riziko.
 (b) (2b) Cílem je učinit binární rozhodnutí Y (*yes*, houba je akceptována) nebo N (*no*, houba se vyhodí). Penalizační funkce je:

$$\begin{aligned} W(E, Y) &= 0, & W(E, N) &= 1, \\ W(T, Y) &= 1000, & W(T, N) &= 0. \end{aligned}$$

Určete bayesovskou optimální strategii, tedy optimální rozhodnutí pro všechny výstupy systému (0, 1, 2).

- (c) (2b) Popište krátce formálně (pomocí vzorců) jak se optimální bayesovská strategie spočítá.
 2. (6b) Odhad parametrů.

Kostkou se hází opakovaně tolíkrát, dokud se na horní straně neobjeví „šestka“. Pravděpodobnost, že se to stane při n -tém hodu, je

$$p(n) = (1 - q)^{n-1}q, \quad (2)$$

kde q je pravděpodonost, že na kostce padne „šestka“.

- (a) (1b) Zdůvodněte správnost vzorce pro $p(n)$.
 (b) (3b) Jaký je maximálně vědohodný odhad pro q pokud šestka padla v k -té hodu?
 (c) (2b) Předpokládejme, že kostkou hodíme maximálně n_{max} -krát. Jaký je maximálně věrohodný odhad pro q , když šestka ani v posledním hodu nepadne?
3. (4b) Neuronové sítě.
- (a) (1b) Vysvětlete, co je vanishing gradient a jak se tento problém řeší.
 (b) (1b) Co je max-pooling? Proč se používá?
 (c) (1b) Co je softmax?
 (d) (1b) Co je metoda stochastic gradient descent?
4. (4b) Vysvětlete, co je K-means++. Co přináší použití K-means++ ve srovnání se standardním K-means?