Machine Learning – June 23, 2020

Time limit: **2 hours 15 minutes**.

EXERCISE 1

Consider a setting where the input space I is the set of finite strings over the characters $a, b, c, \cdots, z$. Notice that input strings can be of different length.

Given the following dataset D:

x	**t**
a	1
ab	1
$caza$	4
$ayka$	4
$aabba$	9
aaa	9
$zazaa$	9
$accaaca$	16
$khaaala$	16
$akdfkkatyuakka$	16
$jhxaaksrtkaatyuap$	25

1. Identify the learning problem at hand, in particular the form of the target function, and define a suitable linear model for it.
2. Apply the kernel trick to the model defined above and provide the analytical form of the corresponding error function.
3. Define the solution obtained with your choices for the dataset D.

EXERCISE 2

A secret string $s = b_0b_1b_2b_3$ of 4 bits fulfills the following constraints:

- if $b_0 = 0$, s contains an even number of 0's and 1's;
- if $b_0 = 1$, s contains at least three 1's.

No other prior information about s is available.

1. Define the prior probability distribution $P(s)$ of the hypothesis string s.
2. Assuming $b_0 = 0$, define the conditional probability distribution $P(s|b_0 = 0)$ and indicate all maximum a-posteriori hypotheses.
3. Assuming $b_0 = 1$ and $b_1 = 1$, indicate all maximum likelihood hypotheses and compute the likelihood that $b_2 = 1$.

EXERCISE 3

Consider a dataset D for the binary classification problem $f : \Re^3 \mapsto \{A, B\}$.

1. Describe a probabilistic generative model for such a classification problem, assuming Gaussian distributions.
2. Identify the parameters of the model and determine the size of the model (i.e., the number of independent parameters).

EXERCISE 4

1. Describe the k-armed bandit problem (also known as One-state MDP).
2. Describe the Reinforcement Learning procedure to compute the optimal policy in the k-armed bandit problem with stochastic behavior and unknown functions.

EXERCISE 5

Consider that the output of layer l of a CNN is the set of feature maps M with size $256 \times 256 \times 64$

1. What is the size of the feature maps N when max-pooling with a 2×2 kernel and stride 2 is applied on M?
2. Design a convolutional layer which, when applied on M, produces feature maps with the same size as N. Describe all the relevant parameters of the layer you have designed.
3. What happens if the non-linear activation functions of the hidden layers of the CNN are replaced with linear functions? Is the effective depth of the network affected and how?

EXERCISE 6

Consider N convolutional neural networks trained to classify images of cats and dogs with a corresponding confidence value (output of sigmoid activation function)

- Describe a way to combine the predictions of the CNNs in order to get a single more accurate prediction.
- Assume $N = 3$, class '0' represents dogs, class '1' cats and for a given image the three CNN outputs are: $(0.912, 0.432, 0.444)$. Apply the method described above to classify the image using the predictions of the three CNNs.