Machine Learning II Online Klausur Sommersemester 2021

Multiple Choice:

Which of the following is True: In a Hidden Markov model (HMM):

- ○ a. $p(q^t = S_i \mid q^{t-2} = S_k) = p(q^t = S_i \mid q^{t-1} = S_j, q^{t-3} = S_l)$.
- ○ b. $p(q^t = S_i \mid q^{t-2} = S_k) = p(q^t = S_i \mid q^{t-1} = S_j, q^{t-2} = S_k)$.
- ◉ c. $p(q^t = S_i \mid q^{t-2} = S_k) = p(q^t = S_i \mid q^{t-1} = S_j)$.
- ○ d. $p(q^t = S_i \mid q^{t-2} = S_k) = p(q^t = S_i \mid q^{t-2} = S_k, q^{t-3} = S_l)$.

Which of the following is True: Canonical Correlation Analysis (CCA):

- ◉ a. Finds which projections of two multivariate random variables are maximally correlated.
- ○ b. Finds which dimensions of two multivariate random variables are maximally correlated.
- ○ c. Finds which projections of a multivariate random variable are maximally correlated.
- ○ d. Finds which dimensions of a multivariate random variable are maximally correlated.

Which of the following is True: The (negative) energy of a conditional RBM is (up to a constant and/or factor):

- ◉ a. The log-probability associated to a state $(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{y})$.
- ○ b. The probability associated to a state $(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{y})$.
- ○ c. The probability associated to an input observation $\boldsymbol{x}$ and output $\boldsymbol{y}$.
- ○ d. The log-probability associated to an input observation $\boldsymbol{x}$ and output $\boldsymbol{y}$.

Which of the following is True: The hard-margin SVDD seeks to:

- ○ a. Find the minimum enclosing sphere that contains all outlier data points in it.
- ○ b. Find the maximal radius sphere that contains only outlier data in it.
- ◉ c. Find the minimum enclosing sphere that contains all non-outlier data points in it.
- ○ d. Find the maximal radius sphere that contains no outlier data in it.

Applications:

At a large trade fair, 1500 companies are presenting. Each company that presents is associated with a number of categories (e.g. manufacturing, systems, consulting, etc.). There are 30 such categories. Conference visitors (65000 registrants) were also asked a few weeks before the event, for the categories they are the most interested in. The trade fair takes place in a large hall. The trade fair organizers would like to place the companies exhibition booths in the hall in a way that reduces congestion. More precisely, they would like to assign for each company a reasonable spatial coordinate in the hall at which the booth should be.

Select a combination of methods that can be used to address the problem above.

- (selected) a. t-SNE + structured kernels
- b. One-class SVM + structured kernels
- c. neural networks + Explainable AI
- d. neural networks + Hidden Markov models

Describe for this particular application how the data should be processed (e.g. which features are extracted), and if applicable, how to build a representation of the data or some similarity model.

Explain how the data and its representation/similarity model, is fed to the machine learning algorithm, how to set the parameters of the algorithm for the purpose of the application, and what the machine learning algorithm optimizes.

Explain how the output of the learning algorithm or analysis is used concretely to solve the application problem above.

Kernel Methods:

A kernel is positive semi-definite, if it satisfies the property

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j k(x_i, x_j) \geq 0$$

for all inputs $x_1, \ldots, x_N$ and choice of real numbers $c_1, \ldots, c_N$.

Let $\mathcal{T}$ and $\mathcal{T}'$ be two documents represented by the set of words composing them. Let $\mathcal{W}$ denotes the set of all words in the dictionary. We consider the structured kernel

$$k(\mathcal{T}, \mathcal{T}') = \Big(\sum_{a \in \mathcal{T}} \sum_{b \in \mathcal{T}'} I(a = b)\Big) - \theta$$

where I is an indicator function and where we assume $\theta \geq 0$.

Considering that the input domain is restricted to email documents which always include the words $\{\text{from}, \text{to}, \text{subject}\}$, give the maximum value of the parameter θ which ensures that the kernel remains positive semi-definite.

Give for the special case $\theta = 0$ a feature map $\phi(\mathcal{T})$ associated to this kernel.

Give for the general case θ, and incorporating the previously stated constraints on the domain, a feature map $\phi(\mathcal{T})$ associated to this kernel. (Note: the feature map must be a vector of real-valued numbers, in particular, it cannot incorporate complex numbers).

Compute $\|\phi(\mathcal{T})\|$ where $\mathcal{T}$ is a document taken from the domain stated above and in addition to words that are included in every document of the domain, also includes the additional sentence "for more detailed information we refer to the appendix".

Conditional Restricted Boltzmann Machines:

We consider a conditional RBM (CRBM) model composed of one input unit $x \in \{0,1\}$, one hidden unit $h \in \{0,1\}$ and one output unit $y \in \{0,1\}$. The CRBM assigns to each joint configuration of units the probability:

$$P(x,h,y) = \frac{1}{Z}\exp(-E(x,h,y))$$

where E is an energy function to be specified and where Z is a normalization constant. In the following, we will be interested in the marginalized distribution $P(x,y)$ for the different states $(x,y) \in \{0,1\}^2$.

We consider the energy function:

$$E(x,h,y) = -xwh + hwy$$

where $w \in \mathbb{R}$ is a parameter of the model. Using the equations above, compute the probability function $P(x,y)$ for each state $(x,y) \in \{0,1\}^2$.

We now consider the "structured output" scenario where we would like to learn the parameter w such that we can predict y from x. Compute the conditional probability $p(y|x)$ for all cases $(x,y) \in \{0,1\}^2$.

Explain how the parameter w controls the correlation between the input x and the output of the model y.

Write the free energy function associated to the CRBM defined above.

Programming:

We consider a Hidden Markov Model defined by the transition matrix A of size $h \times h$, the emission matrix B of size $h \times d$, and the initial state vector π of size h. All these matrices are stored in numpy arrays. The number of states h is a hyperparameter to be selected, and the number of dimensions d is problem dependent.

We would like to build a function that checks whether the HMM given as input is a valid HMM (i.e. matrices have correct dimensions, and entries of the matrices correspond to valid probability distributions).

Implement such function for the case where the HMM models sequences of nucleotides. (In such case, we choose the set of observed symbols to be the four bases GATC, therefore, $d = 4$.)

Create a method that simulates the HMM described above for a given number of iterations T, and as a result returns a sequence of observed symbols.