DIPARTIMENTO DI ELETTRONICA,
INFORMAZIONE E BIOINGEGNERIA

Politecnico di Milano

Machine Learning (Code: 097683)
September 11, 2018

Name:
Surname:
Student ID:
Row: Column:

Time: 2 hours 30 minutes Prof. Marcello Restelli Maximum Marks: 34

- The following exam is composed of **10 exercises** (one per page). The first page needs to be filled with your **name, surname and student ID**. The following pages should be used **only in the large squares** present on each page. Any solution provided either outside these spaces or **without a motivation** will not be considered for the final mark.
- During this exam you are **not allowed to use electronic devices** like laptops, smartphones, tablets and/or similar. As well, you are not allowed to bring with you any kind of note, book, written scheme and/or similar. You are also not allowed to communicate with other students during the exam.
- The first reported violation of the above mentioned rules will be annotated on the exam and will be considered for the final mark decision. The second reported violation of the above mentioned rules will imply the immediate expulsion of the student from the exam room and the **annulment of the exam**.
- You are allowed to write the exam either with a pen (black or blue) or a pencil. It is your responsibility to provide a readable solution. We will not be held accountable for accidental partial or total cancellation of the exam.
- The exam can be written either in **English** or **Italian**.
- You are allowed to withdraw from the exam at any time without any penalty. You are allowed to leave the room not early than half the time of the duration of the exam. You are not allowed to keep the text of the exam with you while leaving the room.
- **Three of the points will be given on the basis on how quick you are in solving the exam. If you finish earlier than 45 min before the end of the exam you will get 3 points, if you finish earlier than 30 min you will get 2 points and if you finish earlier than 15 min you will get 1 point (the points cannot be accumulated).**

Ex. 1	Ex. 2	Ex. 3	Ex. 4	Ex. 5	Ex. 6	Ex. 7	Ex. 8	Ex. 9	Ex. 10	Time	Tot.
/ 5	/ 5	/ 5	/ 2	/ 2	/ 2	/ 2	/ 2	/ 3	/ 3	/ 3	/ 34

Exercise 1 (5 marks)

Describe and compare the ridge regression and the LASSO algorithms.

Exercise 2 (5 marks)

Describe the Principal Component Analysis technique.

Exercise 3 (5 marks)

Describe and compare Value Iteration and Policy Iteration algorithms.

Exercise 4 (2 marks)

Consider the following code lines in `MATLAB`:

```
 1 for ii = 1:n %Initialization
 2   beta_dist(ii) = makedist('Beta','a',1,'b',1);
 3 end
 4 for tt = 1:T %Main loop
 5   for ii = 1:n
 6     hat_r(ii) = beta_dist(ii).random();
 7   end
 8   [~, index] = max(hat_r);
 9   for ii = 1:n
10     outcome = reward(index).random();
11   end
12   beta_dist(index).a = beta_dist(index).a + outcome;
13   beta_dist(index).b = beta_dist(index).b + 1 - outcome;
14 end
```

Tell which algorithm is implemented by the previous code. Is it sound? Are there some mistakes? Enunciate all the assumptions on the environment that are required to use the algorithm implemented in the snippet.

The snippet implements the Thompson Sampling algorithm, an algorithm used to manage the stochastic MAB problem. It requires the initialization of prior distributions for the rewards of each arm (Lines $1-3$), which in this case are uniform distributions. At each round the algorithm samples from each distribution (Lines $5-7$) and selects the one providing the largest sample (Line 8). After that it receives the reward from this arm and updates its prior (Lines $9-13$). The loop in Lines $9-11$ is unnecessary but does not hurt the execution of the algorithm.

The assumption we require s.t. the implemented algorithm is sound are the customary for a stochastic MAB problem. In the specific, since we are updating the Beta prior, we are assuming that the reward has a Bernoulli distribution, otherwise the update would not be correctly executed.

Exercise 5 (2 marks)

Categorize the following ML problems:

- Learn how to play StarCraft (strategic videogame) from recorded games;
- Design an automatic medical system providing a diagnosis based on symptoms;
- Determine which are the most important factors (age, gender, geographical position) when trying to predict what a user is likely to buy;
- Determine the estimated time of arrival of a train.

Provide motivations for your answers.

- Starcraft: The problem is the problem of off-policy control on an MDP, therefore it is an RL problem;
- Medical system: we want to determine which kind of illness, among a finite set, a patient has based on the symptoms, which are the features of a patient. This scenario fits perfectly the classification problem;
- Most important factors: we would like to understand which ones are the features that mostly impact on a problem, therefore, we are facing a feature selection/extraction problem;
- Arrival of trains: in this problem the output variable to be predicted assumes ordered values. If we are given historical data and features important to predict the arrival time (e.g., meteorological information, train status) this is a regression problem, otherwise it is a physics problem (i.e., given a speed profile and a distance determine the time to cover the distance).

Exercise 6 (2 marks)

After performing Linear regression on a dataset, we get eigenvalues $\{0.00000000178,\ 0.014,\ 12\}$ for the matrix $(\Phi^T\Phi)$. Assume we want to perform Ridge regression. What would be meaningful values for the λ regularization coefficient of the matrix $(\Phi^T\Phi + \lambda I)$? Motivate your answer.

The aim of ridge regression is to regularize the design matrix of linear regression. To do so we need add to $\Phi^T\Phi$ an identity matrix λI s.t. the eigenvalues of the resulting matrix are larger than λ. Therefore, values for lambda $\lambda < 0.00000000178$ does not accomplish the task of regularizing the design matrix.

On the other hand, one is not allowed to choose values which are too large for the parameter λ, otherwise the information provided by the data would be disregarded by the regressor. To choose the suitable value for the parameter one might use crossvalidation techniques, evaluating different values for the parameter λ, e.g., $\lambda \in (0.00000000178, 0.014)$.

Exercise 7 (2 marks)

Tell if the following statements are true or false. Motivate your answers.

1. There are multiple optimal policies in an MDP;
2. In finite time MDPs, one should consider only stationary optimal policies;
3. The value function $V^{\pi^*}(s)$ contains all the information required to apply the optimal policy π^* on a given MDP;
4. There are multiple optimal value functions in an MDP.

1. TRUE: there might be the case that two different action provide the same expected reward in a specific state, and, therefore, two different policies are optimal;
2. FALSE: if the time horizon in finite the optimal policy might change over time as the end of the episode approaches. One might consider stationary policies if she introduces the time as part of the state;
3. FALSE: there is no information about the action used to achieve the cumulative expected reward of $V^{\pi^*}(s)$ in each state;
4. FALSE: the optimal value function is the unique fixed point of the Bellman optimality equation.

Exercise 8 (2 marks)

Tell if the following statements about the perceptron algorithm for classification are true or false.

1. We are guaranteed that the loss of the processed datum decreases as we apply the perceptron update step;
2. Shuffling the initial data is fundamental for the perceptron optimization procedure;
3. The solution of the Logistic regression and the one of the perceptron always coincide;
4. There exists a unique solution to the minimization of the perceptron loss if the data are linearly separable.

1. TRUE: we are guaranteed that the update of the perceptron is improving the performance of the learner on the specific sample we are processing. More precisely, we are guaranteed that the loss does not increase on the datum;
2. TRUE/FALSE: since the optimization algorithm of the perceptron is an online procedure, the order in which the data are provided to the algorithm might change the solution the learner achieves, as well as the speed of the learning process;
3. FALSE: even if the two methods uses the same optimization algorithm, they have different loss functions and therefore their solution might be different;
4. FALSE: if the data are linearly separable there might exist multiple lines achieving a perfect classification of the dataset, which are all feasible solutions provided by the perceptron.

Exercise 9 (3 marks)

Show that the VC dimension of a closed interval $[a, b]$ on $\mathbb{R}$ is 2. Provide a PAC bound with confidence at least $1 - \delta = 1 - 4e^{-7}$ for the previous concept when we have $N = \lfloor e^{11} - 1 \rfloor$ samples and an error on the training set of $L_{train}(h) = \frac{1}{e^{10}}$.

Hints: use the fact that $e^{k-1} \leq \lfloor e^k - 1 \rfloor \leq e^k - 1$.

To show that the classifier H has $VC(H) = 2$ we start by showing that $VC(H) \geq 2$ and then that $VC(H) < 3$. By enumeration (pick any two points $x_1 < x_2$ on the real line) we have that an interval $[a, b]$ can separate any partition of these points, therefore $VC(H) \geq 2$. Select then 3 points on the real line x_1, x_2 and x_3. If two of them coincides you are not able to shatter them. Conversely assume that they are distinct and without loss of generality $x_1 < x_2 < x_3$. If we assign a alternate labels to these points we are not able to shatter them. Thus, $VC(H) < 3$, which concludes the proof.

The PAC bound for infinite continuous hypothesis space is:

$$L_{true} \leq L_{train} + \sqrt{\frac{VC(H)\left(\ln(\frac{2N}{VC(H)}) + 1\right) + \ln\frac{4}{\delta}}{N}}$$

Substituting we have:

$$L_{true} \leq \frac{1}{e^{10}} + \sqrt{\frac{2(\ln(\lfloor e^{11} - 1 \rfloor) + 1) + 7}{\lfloor e^{11} - 1 \rfloor}} \leq \frac{1}{e^{10}} + \sqrt{\frac{31}{e^{10}}}$$

Exercise 10 (3 marks)

Consider an MDP with three states $\{s_1, s_2, s_3\}$ and actions $\{d, so, em\}$. Applying the policy π we have the following action-value function $Q(s, a)$ for the three states and the three actions when we consider different discount factors γ:

	$\gamma = 0.9$			$\gamma = 0.95$			$\gamma = 0.99$		
	d	so	cm	d	so	cm	d	so	cm
s_1	35	25	0	95	90	0	780	785	0
s_2	55	0	45	120	0	125	810	0	825
s_3	165	0	0	240	0	0	940	0	0

- Provide the optimal policy π^* for each discount factor γ.
- What is the expected reward for π^* if the initial state distribution is $(0.4, 0.4, 0.2)$.
- Which γ would you choose for this specific problem?

Assuming that the policy π is explorative enough (visits all the states and tries all the available actions) the optimal policy is the one which maximize the value function in each state. Therefore:

$$\pi^*(0.9) = (d,\ d,\ d),$$
$$\pi^*(0.95) = (d,\ cm,\ d),$$
$$\pi^*(0.99) = (so,\ cm,\ d).$$

The expected reward $R(\gamma)$ given an initial distribution $\rho = (0.4, 0.4, 0.2)$ is given by $\rho^T V^*(\gamma)$, therefore:

$$R(0.9) = (0.4, 0.4, 0.2)^T (35,\ 55,\ 165) = 14 + 22 + 33 = 69,$$
$$R(0.95) = (0.4, 0.4, 0.2)^T (95,\ 125,\ 240) = 38 + 50 + 48 = 136,$$
$$R(0.99) = (0.4, 0.4, 0.2)^T (785,\ 825,\ 940) = 314 + 330 + 188 = 832.$$

The third question does not make sense, since the discount factor γ is not a parameter that should be chosen by the learner, but a characteristic of the MDP. The use of different values of γ depends on the fact that the problem requires to be more far-sighted or myopic.