DIPARTIMENTO DI ELETTRONICA,
INFORMAZIONE E BIOINGEGNERIA
Politecnico di Milano
Machine Learning (Code: 097683)
January 23, 2019

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 the two problems tackled by Reinforcement Learning (RL): prediction and control. Describe how the Monte Carlo RL technique can be used to solve these two problems.

Exercise 2 (5 marks)

Describe the techniques called *Ridge regression* and *Lasso regression*. Compare them with respect to their properties.

Exercise 3 (5 marks)

Describe the purpose of using *kernels* in Machine Learning techniques. How can you construct a valid Kernel? Provide an example of a ML method using kernels and describe the specific advantage of using them for this method.

Exercise 4 (2 marks)

Consider the following code lines in MATLAB and a dataset with input X and output Y:

```
P = 42;
n_data = size(X,1);
for order = 0:9
  MSE(order + 1) = 0;
  for pp = 1:P
    idx = 1 + round(n_data*(pp-1)/P) : round(n_data*pp/P);
    x_trai = X; x_trai(idx,:) = [];
    y_trai = Y; y_trai(idx,:) = [];
    x_vali = X(idx,:); y_vali = Y(idx,:);
    LM = fitlm(x_trai, y_trai, ['poly' num2str(order)]);
    y_pred = predict(LM, x_vali);
    MSE(order+1) = MSE(order+1) + mean((y_pred - y_vali).^2);
  end
  MSE(order+1) = MSE(order+1) / P;
end
```

Tell which procedure is implemented by the previous code. Is it sound? Are there some mistakes? Which values are feasible for the variable P? Is P = n_data a feasible value?

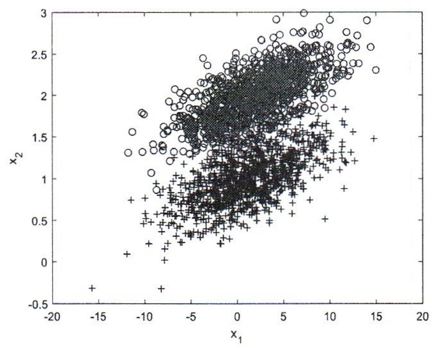
Exercise 5 (2 marks)

Tell if the following statements are true or false. Provide adequate motivations to your answer.

1. Reinforcement Learning (RL) techniques use a tabular representation of MDPs to handle continuous state and/or action spaces;
2. We can use data coming from sub-optimal policies to learn the optimal one;
3. In RL we always estimate the model of the environment;
4. In RL we require to have the model of the environment.

Exercise 6 (2 marks)

Consider the following dataset:



1. Draw the direction of the principal components (on the figure, labeling them as the first and second one) and provide an approximate and consistent guess of the values of the loadings. Are they unique? Why?
2. Is the first principal component a good feature to train a model to classify between the two classes (crosses and circles)? Provide a motivation for your answer.

Exercise 7 (2 marks)

Consider separately the following characteristics for an ML problem:

1. Small dataset;
2. Limited computational resources for training;
3. Limited computational resources for prediction;
4. Prior information on data distribution.

Provide motivations for the use of either a parametric or non-parametric method.

Exercise 8 (2 marks)

Categorize the following ML problems:

1. Pricing goods for an e-commerce website;
2. Teaching a robot to play table tennis;
3. Predicting housing prices for real estate;
4. Identifying counterfeit notes and coins.

Exercise 9 (3 marks)

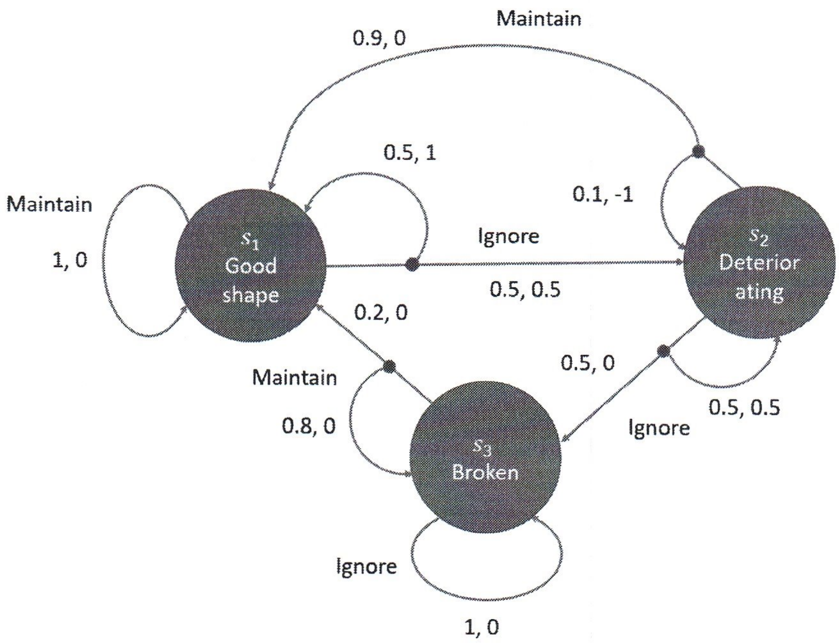
Consider the Thompson Sampling algorithm. Assume to have the following posterior distributions $Beta_i(\alpha_t, \beta_t)$ for arms $\mathcal{A} = \{a_1, \ldots, a_5\}$ rewards, which are distributed as Bernoulli random variables with mean μ_i, and you extracted from them the samples $\hat{r}(a_i)$:

$$
\begin{array}{llllll}
a_1: & \alpha_t = 1 & \beta_t = 5 & \hat{r}(a_1) = 0.63 & \mu_1 = 0.1 \\
a_2: & \alpha_t = 6 & \beta_t = 4 & \hat{r}(a_2) = 0.35 & \mu_2 = 0.5 \\
a_3: & \alpha_t = 11 & \beta_t = 23 & \hat{r}(a_3) = 0.16 & \mu_3 = 0.3 \\
a_4: & \alpha_t = 12 & \beta_t = 25 & \hat{r}(a_4) = 0.22 & \mu_4 = 0.2 \\
a_5: & \alpha_t = 38 & \beta_t = 21 & \hat{r}(a_5) = 0.7 & \mu_5 = 0.6
\end{array}
$$

1. How much pseudo-regret the TS algorithm accumulated so far, assuming we started from uniform $Beta(1, 1)$ priors?
2. Which one of the previous posteriors is the most peaked one?
3. What would UCB1 have chosen for the next round? Assume Bernoulli rewards and that in the Bayesian setting we started from uniform $Beta(1, 1)$ priors?

Exercise 10 (3 marks)

Consider the following MDP modeling the state of a machinery (near the state transitions you have the pair *probability, reward*):



1. Provide the transition matrix for the policy $\pi(I|s_1) = 1, \pi(M|s_2) = 1, \pi(M|s_3) = 1$;
2. Compute the value function for the previous policy in the case the MDP stops after two steps;
3. Compute the action-value function for each state-action pair in the case the MDP stops after a single step.