EFREI Paris

Machine Learning - Rattrapage S7 Exam 2022-04-30

Personal Data

Family Name:
Given Name:
Signature: checked

Registration Number

0	☐	☐	☐	☐	☐	☐	☐	☐	0
1	☐	☐	☐	☐	☐	☐	☐	☐	1
2	☐	☐	☐	☐	☐	☐	☐	☐	2
3	☐	☐	☐	☐	☐	☐	☐	☐	3
4	☐	☐	☐	☐	☐	☐	☐	☐	4
5	☐	☐	☐	☐	☐	☐	☐	☐	5
6	☐	☐	☐	☐	☐	☐	☐	☐	6
7	☐	☐	☐	☐	☐	☐	☐	☐	7
8	☐	☐	☐	☐	☐	☐	☐	☐	8
9	☐	☐	☐	☐	☐	☐	☐	☐	9

In this section **no** changes or modifications must be made!

Type	Exam ID(Machine Learning - Rattrapage)
120	22043000001

Please mark the boxes carefully: ☒ Not marked: ☐ or ■

This document is scanned automatically. Please keep clean and do not bend or fold. For filling in the document please use a **blue or black pen**.

Only clearly marked and positionally accurate crosses will be processed!

Answers 1 - 15

	a	b	c	d	e
1	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	
3	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	
5	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	
7	☐	☐	☐	☐	
8	☐	☐	☐	☐	
9	☐	☐	☐	☐	
10	☐	☐	☐	☐	☐
11	☐	☐	☐	☐	☐
12	☐	☐	☐	☐	
13	☐	☐	☐		
14	☐	☐	☐	☐	☐
15	☐	☐	☐	☐	
	a	b	c	d	e

Answers 16 - 19

	a	b	c	d
16	☐	☐	☐	☐
17	☐	☐	☐	☐
18	☐	☐	☐	☐
19	☐	☐	☐	☐
	a	b	c	d

1. You're running a company and you want to develop learning algorithms to address each of three problems.

   Problem 1: you have a large inventory of identical items. You want to predict how many of these items you sell within the next three months.

   Problem 2: You'd like to write software to examine each individual of your customer's accounts, and for each account decide whether or not the account has been hacked or compromised.

   Problem 3: You want to launch a marketing campaign for a product that you want to sell. So you want to group your customers with respect of their information to be able to target the customers who may be interested in the product.

   So, for each of these problems, should they be treated as a classification problem, as a regression problem or a clustering problem?

   (a) Treat problem 1 as a regression problem, problem 2 as a classification problem, problem 3 as clustering problem.
   (b) Treat problem 1 as a classification problem, problem 2 as a clustering problem, problem 3 as regression problem.
   (c) Treat all as regression problems.
   (d) Treat all as classification problems
   (e) Treat problem 1 as a classification problem, problem 2 as a classification problem, problem 3 as clustering problem.

2. Which of the following problems would you treat as **regression** problems?

   (a) An online banking service must be able to determine whether or not a transaction being performed on the site is fraudulent, on the basis of the user's IP address, past transaction history, and so forth.
   (b) You are treating some images in orders to learn whether a cat appears on an image or not.
   (c) The amount of rain that falls in a day is usually measured in either millimeters (mm) or inches. Suppose you use a learning algorithm to predict how much rain will fall tomorrow.
   (d) Suppose you have data about a company employees and you want to build a model that learns the salary of an employee with respect to his age and experience.

3. To estimate the coefficients of a simple linear regression model we minimize the mean squared error (MSE).

   Select the right answers.

   (a) MSE is the mean of the distances between the input variable and the target variable.
   (b) MSE is the mean of the distances between the response values and the target values.
   (c) MSE is the mean of the distances between the response values and the fitted values.
   (d) MSE is the mean of the distances between the response values and the input values.
   (e) MSE is the mean of the distances between the target values and the fitted values.

4. Suppose you have $n = 32$ samples with $p = 7$ features. The formula for estimating the coefficients of the linear model using *Normal Equation* is given by (in matrix form) $\hat{\beta} = (X^T X)^{-1} X^T Y$

   For the given values of n and p. What are the dimensions of $\hat{\beta}, X$ and Y in this equation?

   (a) X is 32×8, Y is 32×1, $\hat{\beta}$ is 8×1
   (b) X is 33×7, Y is 33×7, $\hat{\beta}$ is 7×1
   (c) X is 32×8, Y is 32×8, $\hat{\beta}$ is 8×8
   (d) X is 32×7, Y is 32×1, $\hat{\beta}$ is 7×7

5. Let's say we want to fit a simple linear regression model. x is the input variable and y is the output variable. ω is the set of parameters.

   Check the correct statements.

   (a) $J(\omega_0, \omega_1) = \frac{1}{2n}\sum_{i=1}^{n}\left(f_\omega\left(x^{(i)}\right) - y^{(i)}\right)^2$ is not convex, so we use the binary cross entropy cost function instead.
   (b) We seek the parameters that minimize a cost function
   (c) The model is $f_\omega(x) = \dfrac{1}{1+e^{-\omega' x}}$
   (d) The cost function to be minimized is $J(\omega_0, \omega_1) = \frac{1}{2n}\sum_{i=1}^{n}\left(f_\omega\left(x^{(i)}\right) - y^{(i)}\right)^2$
   (e) The model is $f_\omega(x) = \omega_0 + \omega_1 x = \omega' x$

6. Check the correct statements about the Gradient Descent (GD) method.

   (a) If the function we seek to minimize is not convex, the algorithm will diverge
   (b) GD is an iterative algorithm
   (c) GD is an optimization technique to find the minimum of a convex function
   (d) A function f is convex if and only if f' (the first order derivative) is positive

7. Suppose that X is quantitative and Y is a binary variable and we want to fit a model to describe Y in function of X. Why don't we apply a linear regression model?

   (a) Because a linear model would be a sigmoid s-shaped function that will takes values between 0 and 1.
   (b) Because we may have negative probabilities with linear regression.
   (c) Because Linear regression cannot separate the response into two classes.
   (d) Because R^2 may be greater than 1.

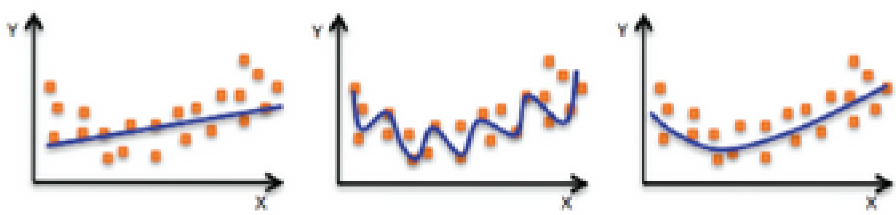
8. To fit a Logistic Regression model, we estimate the coefficients by

   (a) Using the Minimum Likelihood Estimation (MLE).
   (b) Maximizing the log likelihood.
   (c) Using the method of Ordinary Least Square (OLS).
   (d) Minimizing the mean squared error (MSE).

9. Suppose that you have trained a logistic regression classifier, and it outputs on a new example x a prediction $p(x) = 0.65$. This means

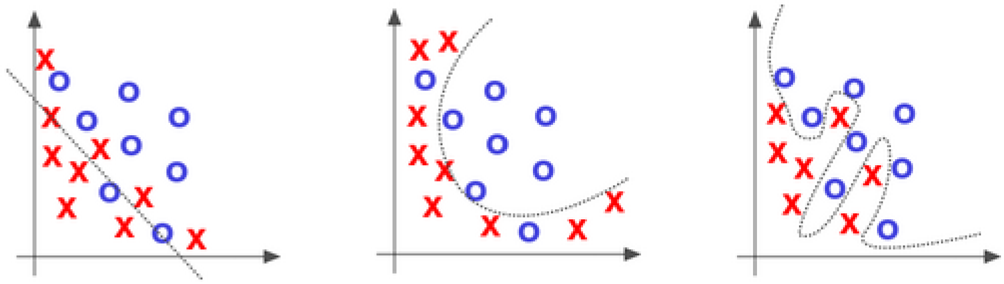
   (a) Our estimate for $P(Y=0|X)$ is 0.65
   (b) Our estimate for $P(Y=1|X)$ is 0.65
   (c) Our estimate for $P(Y=0|X)$ is 0.35
   (d) Our estimate for $P(Y=1|X)$ is 0.35

10. In a simple regression problem, we fitted three different models (model 1 on left, model 2 in the middle, and model 3 on right). Check the correct statements.



   (a) Model 1 is not flexible.
   (b) With model 2, we observe an overfitting.
   (c) The best model between these three is model 2.
   (d) The smallest test error will be obtained with model 2.
   (e) Model 1 underfits the data.

11. In a binary classification problem with two features, we fitted three different models and showed their decision boundaries (model 1 on left, model 2 in the middle, and model 3 on right). Check the correct statements.

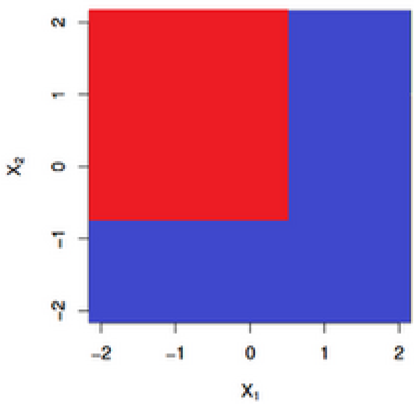


   (a) Model 2 underfits the data.
   (b) The highest accuracy on the train data will be obtained with model 1.
   (c) The decision boundary of model 1 is linear, while the decision boundary of model 2 is quadratic.
   (d) The smallest accuracy on the test data will be obtained with model 3.
   (e) It seems that model 3 is the best one.

12. In Logistic regression, we use the following equation $p(x) = \dfrac{1}{1 + e^{-\omega' x}}$

   (a) $0 \leq \dfrac{p(x)}{1 - p(x)} \leq \infty$.
   (b) $\log\left(\dfrac{p(x)}{1 - p(x)}\right)$ is called the odds and is always between 0 and $+\infty$.
   (c) $\log\left(\dfrac{p(x)}{1 - p(x)}\right)$ is called the logit and can takes any value in $\mathbb{R}$.
   (d) $0 \leq p(x) \leq 1$.

13. Suppose you have a dataset of two features and one binary response variable. The classes are distributed like in following figure (each color is a class). Which classification method seems the best choice in this case?



   (a) Decision tree
   (b) Linear regression
   (c) Logistic regression

14. In a binary classification problem, we fitted a model and ran it on the test set. We obtained the following confusion matrix (rows represent the reality and columns represent the predictions). Check the correct statements.

	0	1
0	57	10
1	7	26

(a) There is 57 True positives.
(b) There is 10 False positives.
(c) There is 7 False positives.
(d) There is 17 bad predictions.
(e) There is 26 False negatives.

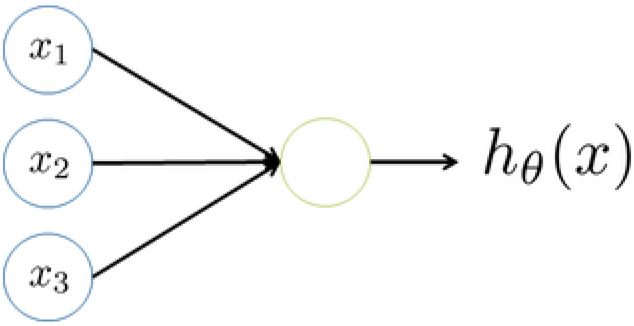
15. In a linear regression problem, we estimated the coefficients ω_j by minimizing the following cost function

$$J(\omega) = \frac{1}{2n}\left[\sum_{i=1}^{n}\left(y^{(i)} - f_\omega\left(x^{(i)}\right)\right)^2 + \lambda\sum_{j=1}^{n}\omega_j^2\right]$$

Check the correct statements.

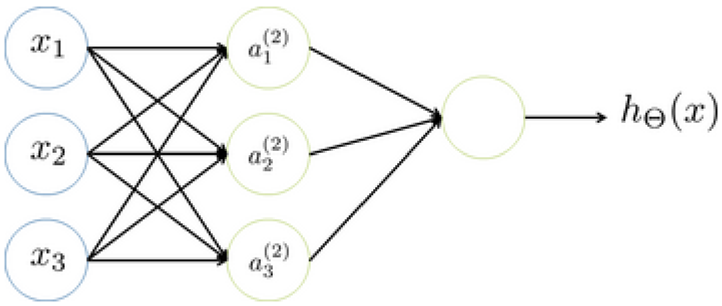
(a) If λ is large, the model will overfit.
(b) This is called Lasso regression.
(c) We can choose the best value of λ by applying a cross-validation.
(d) We use λ to shrink the coefficients to zero.

16. Let the following image of a *Logistic* unit. Check the correct statements.



(a) There is 3 parameters to estimate in this model.
(b) This is equivalent to linear regression.
(c) There is 4 input features here.
(d) We must add the bias unit in the input layer.

17. Let the following image of a simple neural network. Check the correct statements.



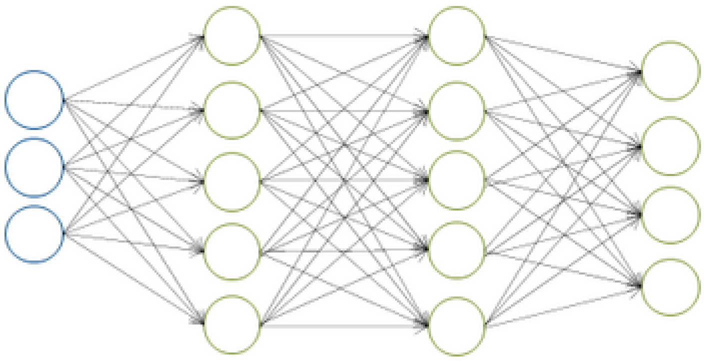
(a) $a_1^{(2)}$ is the activation of unit 1 in layer 2.

(b) $a_3^{(2)}$ is the activation of unit 2 in layer 3.

(c) In this network there is 15 parameters (weights) to estimate.

(d) There is 4 layers in this network.

18. Let the following image of a neural network architecture. Check the correct statements.



(a) In the output layer, $h_\Theta(x) \in \mathbb{R}^4$.

(b) This network architecture is suitable for regression problems.

(c) This network was build to perform multiclass classification.

(d) There is 4 layers in this network.

19. Between the following cost functions, which one is suitable for binary classification neural network? where L is the number of layers and s_l is the number of cells in layer l.

(a) $J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] + \frac{\lambda}{2m}\sum_{j=1}^{p}\theta_j^2$

(b) $J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log\left(h_\Theta\left(x^{(i)}\right)\right)_k + \left(1-y_k^{(i)}\right)\log\left(1-\left(h_\Theta\left(x^{(i)}\right)\right)_k\right)\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(\Theta_{ji}^{(l)}\right)^2$

(c) $J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log\left(h_\Theta\left(x^{(i)}\right)\right)_k + \left(1-y_k^{(i)}\right)\log\left(1-\left(h_\Theta\left(x^{(i)}\right)\right)_k\right)\right]$

(d) $J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(f_\theta\left(x^{(i)}\right)-y^{(i)}\right)^2 + \lambda\sum_{j=1}^{p} w_j^2\right]$