Master in Computer Vision Barcelona

UAB UOC UPC upf

Module: M3. Machine learning for computer vision **Final exam**

Date: February 20th, 2018

- Books, lecture notes, calculators, phones, etc. are not allowed.
- All sheets of paper should have your name.

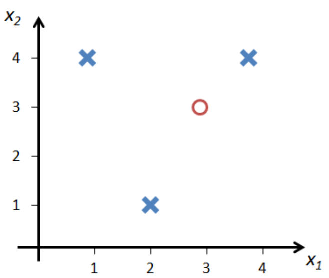
Question 1:

You want to use a learning algorithm to address the following scenarios. How would you treat each of the scenarios: as classification or regression problems?

Scenario	Classification	Regression
Suppose you are working on weather prediction, and you would like to predict whether or not it will be raining at 5pm tomorrow		
Suppose you are working on stock market prediction, and you would like to predict the price of a particular stock tomorrow (measured in dollars)		
Suppose you work in the music industry and you want to develop a service that receives short recordings of music that users have recorded using their mobile phones and responds to each user with the title of the song that was playing		
Suppose you need to decide the budget that Spain should dedicate on research next year. You have data on research spending and high-level academic publications produced of all European Union countries, and you want to decide on the right spending (in EUR) that will make Spain competitive to northern European countries in terms of academic publications.		

Question 2:

Suppose you are fitting a logistic regression classifier: hθ(x) = g(θ0+θ1x1+θ2x2) on the dataset on the right. Which of the following statements are true?



x1	x2	y
1	4	1
2	1	1
3	3	0
4	4	1

	TRUE	FALSE
At the optimal value of θ the value of the cost function will be $J(\theta) \geq 0$		
If we train gradient descent for enough iterations, for some examples $x^{(i)}$ in the training set it is possible to obtain $h_\theta(x^{(i)}) > 1$		
The cost function $J(\theta)$ will be a convex function, so gradient descent should converge to the global minimum		
The positive and negative examples cannot be separated using a straight line. So, gradient descent will fail to converge		
Because the positive and negative examples cannot be separated using a straight line, linear regression will perform as well as logistic regression on this data		

Question 3:

Which of the following statements about regularization are true?

	TRUE	FALSE
Using too large a value of the regularization parameter λ can cause your hypothesis to underfit the data		
Because logistic regression outputs values $0 \le h\theta(x) \le 1$, it's range of output values can only be "shrunk" slightly by regularization anyway, so regularization is generally not helpful for it		
Consider a classification problem. Adding regularization may cause your classifier to incorrectly classify some training examples (which it had correctly classified when not using regularization, i.e. when $\lambda=0$)		
Using a very large value of λ cannot hurt the performance of your hypothesis; the only reason we do not set λ to be too large is to avoid numerical problems		
Because regularization causes the cost function $J(\theta)$ to no longer be convex, gradient descent may not always converge to the global minimum (when $\lambda>0$, and when using an appropriate learning rate α)		

Question 4:

The slack variables are introduced in the SVM

a) To retroproject the features in a higher dimensional space.
b) To solve the quadratic optimisation problem.
c) To increase the maximal margin distance.
d) None of the above.

Question 5:

The solution of the SVM

a) Is a linear combination of all the training samples.
b) Is a non-linear combination of all the training samples.
c) It is a mix between a) and b)
d) None of the above.

Question 6:

For any possible classification problem

a) We can always obtain a zero error in the classification of the validation set by choosing the right classifier.
b) We can always obtain a zero error for the validation set by choosing the right cross validation system.
c) We can always obtain a zero error in the classification of the validation set by choosing the correct hyperparameters.
d) We can always converge to the Bayesian error by choosing the optimal classification system.

Question 7:

With the use of cross validation in an experimental setup

a) We obtain an estimate for the performance of the trained system in terms of average and dispersion.
b) We obtain the best dataset to train our system.
c) We train different ensembles to be used in a boosting approach.
d) None of the above.

Question 8:

In order to solve a non-linear classification problem using a perceptron approach

a) We cannot solve non-linear problems putting perceptrons together.
b) We can add new features as a non-linear combination of the previous features.
c) We can add more perceptrons and connect them following the MLP approach.
d) Both b). and c).

Question 9:

For the back-propagation process used for training a NN

a) It is necessary to derivate the loss function in order to maximize it, by applying the chain rule backwards, from the output layer to the input layer,
b) It is necessary to derivate the loss function in order to minimize it, by applying the chain rule backwards, from the output layer to the input layer,
c) It is necessary to derivate the loss function in order to maximize it, by applying the chain rule backwards, from the input layer to the output layer,
d) It is necessary to derivate the loss function in order to minimize it, by applying the chain rule backwards, from the input layer to the output layer.

Question 10:

The Recurrent Neural Networks RNN

a) Allow the connection with ensambles of other NN to improve performance.
b) Allow the connection with neurons between themselves and in previous layers.
c) Allow the connection with stacked maxpool stages in a recurrent way.
d) None of the above.

Question 11:

The RLU transfer function can be preferred to the sigmoid function in CNNs mainly because

a) It has faster performance, which is relevant to the training phase.
b) It tends to a lower saturation for the neurons output.
c) It has a pure linear behaviour.
d) None of the above.

Question 12:

In CNNs

a) Neurons integrate local information to take profit of the statistical properties of images as spatial redundancy.
b) Neurons integrate global information to take profit of the statistical properties of images as spatial redundancy.
c) Neurons integrate local information to take profit of the statistical properties of images as temporal redundancy.
d) None of the above.

Question 13:

Given the typical architecture of a CNN, we can verify that the perceptive field for a given neuron

a) Is being increased as we advance forward in the different layers.
b) Is being decreased as we advance forward in the different layers.
c) Is kept constant as we advance forward in the different layers.
d) None of the above.

Question 14:

Which are the two main motivations to include dimensionality reduction in feature description? explain

Question 15: (2 points)

Let's assume we have set up a basic BoW framework using SIFT for feature detection and description, K-means for vocabulary construction, hard assignment for image representation and linear SVM for classification.

Propose two possible modifications to this basic framework that could improve the accuracy and/or efficiency of the system. For each modification:

a) Give a brief description about what does it consist of and how should it be integrated into the basic framework.
b) Explain the rationale behind that modification, i.e, why could it improve accuracy or efficiency.

Question 16:

Explain the concept of "simmetry breaking" in the context of deep models' weights initialization.

Question 17:

When training deep neural networks, the learning rate is one of the most difficult hyperparameters to set. Expalin the aim and basic principle behind the adaptive learning rate methods we've seen in the module.

Question 18:

Understanding what has been learned on a trained CNN can be done in different ways:

1. by describing the Neuron Activity of each individual neuron to understand the functionality of each neuron.
2. by inverting feature maps into the image space in order to visualize what is represented in a specific layer, or
3. analyzing whether input modifications allow to visualize if these changes provoke different representations.

Pick one specific approach, briefly specify to which group of methods it belongs to and describe the main ideas of the approach?

Question 19:

According to what we can conclude from the approaches that try to understand CNNs, answer if the following sentences are TRUE or FALSE:

a) The activity of the set of CNN neurons define the high-dimensional space where images are represented through the network.
b) A low correlation between weights of two different neurons means that these two neurons are trained to represent almost the same feature
c) In shallower layers of a CNN each neuron codifies a specific feature. If this feature exist in the image in a specific location, then the response of the neuron at this position is low.
d) CNNs are invertible, since all of the layers involved in typical CNNs (including convolutional, pooling and non-linear layers) are invertible.
e) The top-scoring images are the set of images of the dataset that are better classified by the overall network (with a higher confidence)
f) The main problem of describing the neuron activity by generating a new image that maximizes the activation of a specific neuron is that the result may be a non-realistic image. These approaches require good regularizations.
g) Neurons of shallower layers of a hierarchical CNN are devoted to simple image features
h) Image feature complexity decreases from shallow to deep layers
i) Fully-connected neurons in hierarchical classification architectures are able to codify different versions of the same object.
j) Given the feature maps of a specific image obtained through a trained CNN, we can invert these feature maps towards the image space. This projection shows how the entire input image is described in this layer

Name:____

Answers to test questions. From 4 to 13. Mark the correct answer.

	4	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
a)										
b)										
c)										
d)										

Question 19:

a)	**b)**	**c)**	**d)**	**e)**	**f)**	**g)**	**h)**	**i)**	**j)**

Name:____

Questions 14 and 15

Name:____

Questions 16 and 17

Name:____

Question 18