LELEC2870 LELEC2870 January 2021 exam

Corentin Lingier

TOTAL POINTS

9.8 / 12

QUESTION 1

CNN 1 pts

1.1 layer 1 **0.25 / 0.25**

✓ **- 0 pts** *Correct: convolutional*

1.2 layer 2 **0.25 / 0.25**

✓ **- 0 pts** *Correct: pooling / subsampling*

1.3 layer 3 **0.25 / 0.25**

✓ **- 0 pts** *Correct: fully-connected*

1.4 layer 4 **0.25 / 0.25**

✓ **- 0 pts** *Correct: softmax or Gaussian*

QUESTION 2

Clustering 3 pts

2.1 objective functions **0.55 / 1**

✓ **- 0.45 pts** *Right obj.functions but not comparable: discrete versus continuous; no/wrong discusion as to the similarity between them*

2.2 centroids update **1 / 1**

✓ **- 0 pts** *Correct*

2.3 convergence **0.5 / 1**

✓ **- 0.5 pts** *Missing key elements...*

Systematic decrease for KM? R-M cond. for CL?

QUESTION 3

Model selection 3 pts

3.1 which model selection method **0.4 / 0.5**

✓ **- 0.1 pts** *CV is good but cross-val CV is better: there are advantages and no drawback wrt to CV*

good answer

3.2 justify which model selection method **0.3 / 0.5**

✓ **- 0.2 pts** *vague and/or partly wrong justification*

3.3 numerical score approximates what **0.5 / 0.5**

✓ **- 0 pts** *Correct (including if there was a confusions between generalization and validation error)*

3.4 model selection **0.5 / 0.5**

✓ **- 0 pts** *Correct*

3.5 test set **1 / 1**

✓ **- 0 pts** *Correct*

QUESTION 4

RBFN 3 pts

4.1 nonlinear function **0.5 / 0.5**

✓ **- 0 pts** *Correct*

4.2 learning algorithm 1 / 1

✓ **- 0 pts** *Correct*

4.3 interpretability 0.2 / 0.5

✓ **- 0.3 pts** *not wrong but does not really answer to the question.*

4.4 interpretability fails in HD 0.3 / 0.5

✓ **- 0.2 pts** *not wrong about HD spaces, but does not answer to the question about interpretability of the RBFN*

4.5 methodology with FS 0.3 / 0.5

✓ **- 0.2 pts** *to keep the interpretability of feature we prefer a feature *selection* method, not *extraction/projection* such as PCA, ICA, MDS, SNE,...*

QUESTION 5

5 Data standardization 0.75 / 1

✓ **- 0.25 pts** *Wrong on LR -> Optional (LR is a convex problem and is solvable by inversing a matrix)*

QUESTION 6

6 Grid search & Cross-validation 1 / 1

✓ **- 0 pts** *Found the train/val contamination: Understood that validation data was already in use but that the score was only computed on part of it (in this case 80%)*

Tuning on validation -> understanding seems thus ok

gradescope

First name: Corentin

Family name: Lingier

Noma: 23431700

LELEC2870 : Machine Learning - January 2021 exam

PLEASE READ THIS FIRST!

Organization of the exam:

- For sanitary reasons you will not be allowed to leave your seat during the duration of the exam (2 hours). Even if you have finished please wait until the end.
- At the end of the exam we will ask you
  1. to stop writing (all of you at the same time)
  2. to move row by row, keeping 1m50 distance, and come to the front of the room to give your documents. You have to hand in:
     - **All the enclosed sheets** (including this one), and not any other one (no supplementary sheet, draft sheet, etc.)
     - Arranged in the **correct order** (page 1 first on top, then page 2,...)
     - And ensure that your first name, family name and NOMA are indicated on **every sheet**
- For sanitary reasons we are not allowed to come close to you during the exam to answer to some questions. We are thus sorry that will not be able to answer any question you may have. An exception will be made during the first 10 minutes of the exam: you will have the possibility to ask a question from your seat, while everybody in the room will listen.

Instructions:

- Write **your answers only in the frames**. This gives a good indication of the expected length of your answer. You don't have to use small handwriting to write more! Please do not write anything outside the frames, it will not be considered.
- The exam is open book; this means that **only written notes** are allowed; any electronic device with or without connection to the internet is not allowed.
- Fill in your **first name, name** and **NOMA** on **every page.** Write clearly in capital letters, as this will be interpreted by a computer.

Rating system:

- The project counts for 10 points on 20, including part B of this exam (the part related to the project). Part B of this exam won't alter your project points by more than 2/10. Please prioritize Part A first! Do not spend more than 15-20 minutes on Part B, otherwise you will not have enough time for Part A.
- Part A counts for 10 points on 20.

First name: CORENTIN		**Noma:** 2343 1700
Family name: LINGER		

Part A. Questions on the Course

Question 1 – CNN

Give the different types of layers that are found in a deep convolutional neural network intended to classify images among, say, 10 classes, like Lecun's LeNet 5. List the layer types as they would be stacked in a network, from input to output. (The input image is not considered as a layer. Up to 5 words per layer.)

Layer type 1 =	Convolutional layer
Layer type 2 =	Max pooling layer
Layer type 3 =	Dense (~~pre~~ with flattening)
Layer type 4 =	Gaussian

Question 2 – Clustering

In vector quantization, what is the relationship between K-means (a.k.a. Lloyd's algorithm, LBG, isodata) and competitive learning? What do they have in common? **Discuss briefly**

2.1 whether their objective functions are the same?

The adaptation rule of K-Means is

$$E_{VQ}(Y, X) = \frac{1}{N} \sum \| x_i - y_{j(i)} \|^2$$

where as it is:

$$E_{VQ}(Y, X) = \int \| x - y_{j(x_i)} \|^2 \, p_x(x) \, dx$$ for competitive learning

They both have in common a notion of ~~distance~~ euclidian distance thus they will maintain a codebook based on this metric.

First name: CORENTIN

Noma: 2343 1700

Family name: LINGIER

2.2 how frequently they update the centroids when processing a data set?

K-means:
for K-means, the codebook is adapted after the presentation of the whole dataset

Competitive learning:
~~Only one d~~ One or more centroids are adopted after presenting a data point. Thus the centroids are more often updated in Competitive learning.

2.3 how they will converge depending on a given initialization, and under which conditions?

K-means:
K-means ~~will~~ states that the probability of finding a point on the border of a voroni zone is 0. ~~Thus it~~ And the risk of getting trapped in a local minimum of the error is high

$$E_{VQ} = \frac{1}{N} \sum \|x_i - y_{j(x_i)}\|^2 \text{ with } j(x_i) = \operatorname{argmin}_j \|x_i - y_j\|^2$$

The final quantizer depends on the initial one.

Competitive learning:
Competitive learning ~~may be stuck in~~ gets easily stuck in a local minimum, And given an initialization, some centroids may be lost.

First name: CORENTIN | Noma: 23931700

Family name: LINGIER

Question 3 – Model Selection

You receive a dataset containing 10.000 numerical data in a 20-dimensional space. You hesitate which model you should use (between, for example, a 2-layers MLP with 10 hidden units, a 2-layers MLP with 20 hidden units, a RBFN with 10 Gaussian functions, a deep network, etc.).

3.1 Which of the following methods would you prefer to use in order to decide which model to use: *random choice, validation, cross-validation (CV), K-fold CV, leave-one-out.* **Choose one**:

Cross-validation (CV)

3.2 Please **motivate** your choice:

~~The~~ We are in a data rich situation where we can, using cross-validation assess ~~a~~ ~~st~~ the generalization error. And the models usually take a long time to train thus using k-fold or LOO takes too much time.

3.3 The result of these methods is a *numerical score*. What does this *score* **approximate**? Please use accurate words for your answer:

~~The goodness of fit done on the dataset~~; The generalization error.

3.4 Please **describe** how you would proceed in practice to **select a model** between several possible ones, based on the *score* mentioned in the previous sub-question:

I would make sure to evaluate the models on ~~an~~ unseen dataset and I would take the model with the lowest generalization error

3.5 Please motivate why you need a third set of data, called **test set**, not used in the above methods. Justify why you cannot use the **validation set(s)** used in the CV for the same purpose.

The test set is used to ~~assess~~ estimate the generalized error, and prevent from overfitting while the validation set is used to ~~make~~ take decisions on the model and to tweak its parameters.

First name: CORENTIN	**Noma:** 23431700
Family name: LINGIER	

Question 4 – RBFN

4.1 What is the type of **nonlinear function** that is traditionally used in **RBFN** (Radial-Basis Function Networks)?

Green's function are used in RBFN

4.2 **Describe** shortly the **learning algorithm** of an RBFN

1. Centers c_i: use VQ to find the centers (c_i are centroids).
2. Widths σ_i: use the standard deviation around the centroids to set the widths.
3. Weights w_i: with c_i and σ_i, finding w_i becomes linear. Simply use pseudo-inverse, singular value decomposition or gradient descent.

We could also use gradient descent on all parameters.

4.3 **Explain** why, for a **similar complexity** (same number of layers, same number of units in the hidden layer, ...), the learned parameters of a RBFN might be **more interpretable** than the learned parameters of an MLP.

MLP uses scalar products as arguments for their hidden units. Where RBFN uses euclidean distance which is more interpretable as it is the "natural distance".

First name: CORENTIN Noma: 23931700

Family name: LINGIER

4.4 **Explain** why, on the contrary, the possibility to **interpret the parameters** of a RBFN during and after learning might fail when the data space is very **high-dimensional.**

The euclidean distance in a high-dimensional space doesn't ~~discrimi~~ discriminate well data points. In fact the ratio between a hypersphere and a hypercube as dimension increases, tends to 0 ~~and~~ All data points are casted further aparts And the distance between them tends to the overall expected distance.

4.5 **Explain** how you would proceed to analyse numerical data in the following situation: the regression problem relies on data that are **high-dimensional** (for example 100 features), and **not really numerous** (for example you have 2000 samples). You still want to **use a RBFN model** and **benefit from the interpretation of its parameters**. And at the same time, you would like to be able the **interpret the features** that contribute the most to the prediction. What would be your strategy? You don't have to explain the methods/algorithms that you will use: just mention them (as precisely as possible) and justify your choice(s).

It is one of the reason to use a dimensionality reduction algorithm.
As we want to interpret the ~~results~~ ~~from~~ parameters from the RBFN we should use Dim. Reduction algorithms based on ~~distance~~ euclidean distance:

- METRIC MDS
- Sammon NLM
- CCA

As we want to interpret the features, I would recommend MDS as there is no weighting

First name: CORENTIN	Noma: 23431700
Family name: LINGIER	

Part B. Questions on the Project

Question 5 – Data Standardization

Let us take a dataset with features of different magnitude. For which of the following algorithms is data standardization <u>mandatory</u> (results are different if it isn't applied) OR <u>highly recommended</u> (it may work without it, but it converges faster with) OR <u>optional</u> (it will work the same regardless if it is applied or not). **Put an X in the correct boxes (only one is correct per algorithm!).**

	Mandatory	Highly Recommended	Optional
K Nearest Neighbors	☒	☐	☐
PCA	☒	☐	☐
Linear Regression	☐	☒	☐
Multi-Layer-Perceptron	☐	☒	☐

First name: CORENTIN	**Noma:** 23931700
Family name: LINGIER	

Question 6 – Grid-search & Cross-validation

```python
from sklearn.model_selection import GridSearchCV
from sklearn.neighbors import KNeighborsRegressor as KNN

X_train # Contains training data standardized features
y_train # Contains training data target
score_regression # Fuction given in the project description

params = {
    "n_neighbors": [1, 5, 10, 20]
} # The 'n_neighbors' parameters of the KNN will be set to these different values

gscv = GridSearchCV(
    estimator=KNN(),
    param_grid=params,
    scoring=score_regression,
    cv=5  # 5-fold cross-validation
)

gscv.fit(X_train, y_train)
model = gscv.best_estimator_
"""
From the official documentation:
best_estimator_ : Estimator that was chosen by the search,
                  i.e. estimator which gave highest score
                  (or smallest loss if specified) on the
                  left out data (i.e. the unused fold).
"""

score = score_regression(
    model.predict(X_train),
    y_train
)
print(score)
```

Applying the code snippet (found on the previous page) on the project's dataset, you obtain a score of ≃75%. You conclude that KNN is the best model by a 25% margin compared to your Linear Regressor baseline. The tutor evaluates your test-set and finds a score closer to 50%. **What went wrong during the model evaluation?** Explain your answer in 1-2 sentence(s).

the Score is evaluated ~~on data for which the mode~~ with data that the model has trained on. And ~~ith~~ it has tweaked its parameters to fit the set, thus there is no assessment of the generalized error.