

MLRF Lecture 06

J. Chazalon, LRDE/EPITA, 2021

Agenda for lecture 6

1. Introduction
2. Some classifiers - part 2
3. More theory about classifiers
4. Introduction to practice session 6
5. Course conclusions

Introduction

Lecture 06 part 01

Previously, in MLRF...

Summary of last lecture

Image classification overview

- *Instance vs category* recognition
- Pipeline: feature extraction > encoding > pooling > classification
- Our approach:
input = vectors, output = integers
- Proper evaluation

Classifiers

- Dummy
- kNN
- Distance based
- Generative probabilistic (Naive Bayes...)
- Linear discriminant probabilistic
 - Logistic regression, SVM
 - multiclass

Classifiers (continued)

- Non-linear discriminant probabilistic
 - Data preprocessing
 - Explicit embedding
 - Kernel trick

Evaluation

- Confusion matrix
- {true,false} {positive,negative}
- Accuracy
- How to plot PR curve
- ROC curve
- AUC

Train/validation/test sets separation

Debriefing of practice session 5

Content

Bag of Visual Words classifier

1. Load resources
2. Train a BoVW model
3. Split the dataset into training and validation sets
4. Compute the BoVW descriptor for each image
5. Prepare training structures
6. Train a classifier and evaluate its performance
7. Display some results
8. Test on meme images
9. Compute the results on the test set and export them

Discussion

- Who completed part 1? 2? ...
- Did everyone submit their results?
 - results.json
 - notebook.ipynb
- Any remarks, comments, questions?
- Things to keep, change, remove?

Practice session 5: Take home messages

BoVW is linear classification friendly

Sparsity helps.

- And linear classifiers are to be preferred whenever possible

Data preparation is tedious

- An important part of the time dedicated to data analysis
- Plus we prepared a lot of things for you in the previous sessions!

Scikit-learn is easy and super powerful

- Classifier evaluation in 1 line
- But there is more: parameter tuning, cross-validation, etc. in 1 or 2 lines
- Data preprocessing + classification (pipelines) in 1-3 lines...

Next practice session
Later in part 04