Basics of machine learning
Nicolas Duchateau, Rémi Emonet, Carole Lartizien

Deep Learning for Medical Imaging
Lyon (FR) - 15/04/2019

Program

~30min 1. Introduction
Carole Lartizien

~75min 2. Supervised learning
Rémi Emonet + Carole Lartizien

~75min 3. Unsupervised learning
Nicolas Duchateau + Rémi Emonet

~30min 4. Methods evaluation
Carole Lartizien + Rémi Emonet + Nicolas Duchateau

~30min 5. Conclusion

4. Methods evaluation
Supervised learning in a nutshell

- Split the sample dataset into three parts: a training, a validation and a test dataset
- **Choose a parameterized model function** with parameters $\Theta_1$ and hyperparameters $\Theta_2$ from an hypothesis space $H$
- **Fit the model parameters $\Theta_1$** to the training dataset for a fixed value of $\Theta_2$
  - Choose an error function that measures the misfit between the decision function $D(f(x))$ and the class $y_i$ of all training data points $(x_i, y_i)$
  - Minimize the error function
- Evaluate the performance of your model on the validation dataset
- Retrain your model with another hyperparameter set $\Theta_2$
- Select the best parameter set
- Evaluate the performance of your best model on the test dataset
Objective and definitions

- To estimate the intrinsic performance of a decision model.
- The empirical error on the training dataset is optimistically biased.
- The generalization error is the error achieved on new dataset representative of the data distribution $P$, i.e., data different from the training database.
- How to estimate the generalization error from samples drawn from the data distribution?

Supervised learning

Performance evaluation on a dataset of $N$ samples in a representation space of dimension $d$:

- What performance metric?
- What method to best approximate the generalization performance?

Metrics for the evaluation of supervised classification models

<table>
<thead>
<tr>
<th>Estimated class</th>
<th>True Negative (TN)</th>
<th>False positive (FP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True class</td>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td></td>
<td>False negative (FN)</td>
<td>True Positive (TP)</td>
</tr>
</tbody>
</table>

Confusion matrix

- **Sensitivity** = Fraction of true positives = $TP/(TP+FN)$
- **Specificity** = Fraction of true negatives = $TN/(TN+FP)$
- **Accuracy** = Fraction of correct answers = $(TN+TP)/(TN+TP+FN+FP)$
4. Methods evaluation

Metrics for the evaluation of supervised classification models

\[ AUC = \text{Area under the ROC curve} \]

\[ H(w, b) = \{ x \in \mathbb{R}^d \mid w^T x + b = 0 \} \]

Class 1

\[ H(w, b) = \{ x \in \mathbb{R}^d \mid w^T x + b' = 0 \} \]

Class -1

\[ \text{Sensitivity} = \text{True positives} \]

\[ \text{Seuil} = 1 \]

\[ \text{Seuil} = 0.8 \]

\[ \text{Seuil} = 0.5 \]

\[ \text{Seuil} = 0.3 \]

\[ \text{Seuil} = 0 \]

\[ \text{AUC} = \text{Area under the ROC curve} \]

\[ H(w, b) = \{ x \in \mathbb{R}^d \mid w^T x + b = 0 \} \]

\[ \text{Class} = -1 \]

4. Methods evaluation

Supervised learning

Extension to the multi-class case

\[ \text{Matrice de confusion multi-classes} \]

\[ \text{Precision}(c_j) = \frac{c_{ij}}{\sum_{k} c_{kj}} \]

\[ \text{Rappel}(c_j) = \frac{c_{ij}}{\sum_{j} c_{ij}} \]

\[ \text{Accuracy} = \frac{\sum_{i,j} c_{ij}}{\sum_{i,j} c_{ij}} \]

4. Methods evaluation

Supervised learning

Hold-out

- Principle:
  - Split the dataset into one learning and one validation dataset
  - Learn a model on the training dataset and evaluate its performance on the validation dataset

- Advantages:
  - Easy to implement
  - Low computational cost

- Bottlenecks:
  - Requires a lot of data
  - Sensitive to the database splitting
  - Robustness of the results

\[ \text{Dice}(S, T) = \frac{2 |S \cap T|}{|S| + |T|} \]

\[ D_H(S, T) = \max \left\{ \sup_{s \in S} \inf_{t \in T} d(s, t), \sup_{t \in T} \inf_{s \in S} d(s, t) \right\} \]
4. Methods evaluation

K-folds cross-validation

- Choose a number of folds $k$
- Split the dataset into $k$ subsets $D_1, \ldots, D_k$
- For each subset $k$:
  - Learn the model on the union of $D_j$ subsets with $j \neq i$
  - Evaluate the performances on fold $D_i$
  - Combine performances achieved on all folds $D_j$

if $k = N$ (the number of samples) → leave-one-out (LOO)

[Source: Cours de F. Rossi, Telecom ParisTech]

4. Methods evaluation

Cross-validation

- Advantages:
  - Easy to implement
  - Make use of all available data to estimate the model performance
- Limits:
  - Sensitive to the splitting strategy
  - High computational time
  - Do not provide a model

4. Methods evaluation

Bootstrap

2. Draw one bootstrap dataset $x^b$ by random sampling $N$ samples with replacement from the original dataset $x$
3. Estimate the performance metric by hold-out on the sample $x^b$
4. Estimate the performance metric of the model trained $x^b$ and tested on $x$
5. Compute an estimate of the bias of the hold-out method
6. Evaluate the performance metric by subtracting the estimated bias $w$ from the hold-out performance estimated on the original dataset $x$
4. Methods evaluation

Bootstrap

- Advantages:
  - Easy to implement
  - Make use of all available data to estimate the model performance
  - Provide confidence intervals

- Limits:
  - High computational time
  - Do not provide a model

4. Methods evaluation

Unsupervised...

no “ground truth” labels = not so easy?

→ Some practical examples
  (but highly depends on the data + clinical question)

4. Methods evaluation

Unsupervised

1) Neighborhood graph: short-circuit?

Synthetic data = easy?

Duchateau et al. Med Image Anal 2012

Estimated embeddings

4. Methods evaluation

Unsupervised

1) Neighborhood graph: short-circuit?

Synthetic data = easy?

\[ \epsilon(x_i, x_j) = 1 - \frac{||y_i - y_j||}{d_{KNN}(x_i, x_j)} \]

Error vs. geodesic distance

Duchateau et al. Med Image Anal 2012

Estimated embeddings
4. Methods evaluation

**Unsupervised**

1) Neighborhood graph: short-circuit?

Real data = unknown...

→ Possible option = measure that can assess the presence of a short-circuit

Node flow:

$$F_{node}(x_i) = \sum_{k=1}^{K} F_{edge}(edge_{x_i = min_0(x_i)})$$

Duchateau et al. Med Image Anal 2012

![Node flow images](Duchateau et al. Med Image Anal 2012)

4. Methods evaluation

**Unsupervised**

2) Dimensionality / clusters: On synthetic/clear data, already known

Input space | Ground truth embedding | Synthetic data

Duchateau et al. Med Image Anal 2012
4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

- **Compactness** = data explained by few dimensions?
- **Generalization ability** = error in reconstructing known cases?
- **Specificity** = consistency of generated new data?

**Theoretical error**
ex: error vs. geodesic distance

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

- **Compactness** = data explained by few dimensions?
- **Generalization ability** = error in reconstructing known cases?
- **Specificity** = consistency of generated new data?

**Theoretical error**
ex: error vs. geodesic distance

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

- **Compactness** = data explained by few dimensions?
- **Generalization ability** = error in reconstructing known cases?
- **Specificity** = consistency of generated new data?

**Theoretical error**
ex: error vs. geodesic distance

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

- **Compactness** = data explained by few dimensions?
- **Generalization ability** = error in reconstructing known cases?
- **Specificity** = consistency of generated new data?

**Theoretical error**
ex: error vs. geodesic distance
4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

Consistency of clusters

Cikes et al. Eur J Heart Fail 2019

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

Consistency of clusters

Clinical variables

Cikes et al. Eur J Heart Fail 2019

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

vs. risk stratification:

Cikes et al. Eur J Heart Fail 2019

4. Methods evaluation

Unsupervised

2) Dimensionality / clusters: On real data = ?

vs. risk stratification:

Cikes et al. Eur J Heart Fail 2019
4. Methods evaluation

Unsupervised
3) Linear / non-linear:

vs. physiological consistency:

Linear
Non-linear

Duchateau et al. Med Image Anal 2012

Program

~30min  1. Introduction
         Carole Lartizien

~75min   2. Supervised learning
         Rémi Emonet + Carole Lartizien

~75min   3. Unsupervised learning
         Nicolas Duchateau + Rémi Emonet

~30min   4. Methods evaluation
         Carole Lartizien + Rémi Emonet + Nicolas Duchateau

5. Conclusions / to go further

5. Conclusions

Take-home message

➔ Supervised / unsupervised
   ◆ Can handle complex data
   ◆ Proper definition of your problem vs. application
   ◆ Proper evaluation is required
   ◆ Transparency = open code, open data

➔ For medical imaging applications:
   ◆ "Standard" learning = already powerful to go beyond simple statistics
   ◆ Specificities of medical data vs. computer vision or data science
     ● High dimension
     ● Specific properties (ex: tensor)
5. Conclusions

Take-home message... and challenges

➔ Supervised / unsupervised / semi-supervised
  ➔ Can handle complex data
  ➔ Proper definition of your problem vs. application
  ➔ Proper evaluation is required
  ➔ Transparency = open code, open data
  ➔ Mixing multiple heterogeneous descriptors?

➔ For medical imaging applications:
  ➔ "Standard" learning = already powerful to go beyond simple statistics
  ➔ Specificities of medical data vs. computer vision or data science
    ● High dimension
    ● Specific properties (ex: tensor)
  ➔ Real life data = quality / completeness / amount
  ➔ Physiological prior
  ➔ Consequences of uncertainties?

Thanks!... Any questions?