Multi-target prediction

Outline

- **What is multi-target prediction?**
- **Intuition and Applications**
- **Basic Use-Cases**
- **Multi-target vs. multi-label classification**
- **Different approaches**
- **Predictive Clustering Trees**
- **Measurement of prediction quality**
- **Hands-on demonstration**

What is Multi-target prediction?

- **Multi-target prediction, multi-output or multi-task learning.**
- **The goal is to predict multiple dependent variables (targets) simultaneously from a set of input variables (features).**
- **Particularly relevant in complex real-world scenarios where multiple outcomes are interrelated and can influence each other.**

Intuition

- **Shared Representation: the targets share some common underlying factors. By learning these shared representations, the model can make more informed predictions for each target.**

- **Exploiting Correlations: Multi-target models aim to exploit the correlations and interactions between targets. E.g., in a health-related dataset.**

- **Joint Learning: Where the model learns to optimize the predictions for multiple targets in a coordinated manner. This joint learning approach can help in uncovering insights that may not be apparent when targets are considered in isolation.**

Applications

- **Complex Interdependencies: Traditional single-target prediction models might ignore interdependencies of targets, potentially leading to suboptimal performance.**

- **Efficiency: Reducing the need to train separate models for each target. This can lead to more efficient use of computational resources and data.**

- **Improved Generalization: Multi-target models can potentially improve prediction accuracy by learning shared representations that capture underlying patterns relevant to multiple targets.**

i) Basic Use-Cases: Weather Forecast

- **Applied to many domains, such as genomics, finance, and recommendation systems, where multiple outcomes or variables are of interest are not independent of each other.**

- **An illustrative use cases of multi-target prediction: in environmental modeling, specifically in predicting various aspects of weather or climate conditions from a set of input variables.**

- **Forecast multiple weather parameters, such as temperature, humidity, precipitation, wind speed, and air pressure, from historical and current weather data.**

ii) Basic Use-Cases: Drug Discovery

- **In drug discovery and personalized medicine, it's crucial to understand how different compounds interact with various genetic markers or cellular processes.**

- **A single drug can have multiple effects, impacting various genes, proteins, or pathways.**

- **Traditional approaches might involve building separate models to predict the effect of compounds on each marker or response, which can be inefficient and may ignore the interdependencies between these effects.**

Multi-target vs. multi-label prediction

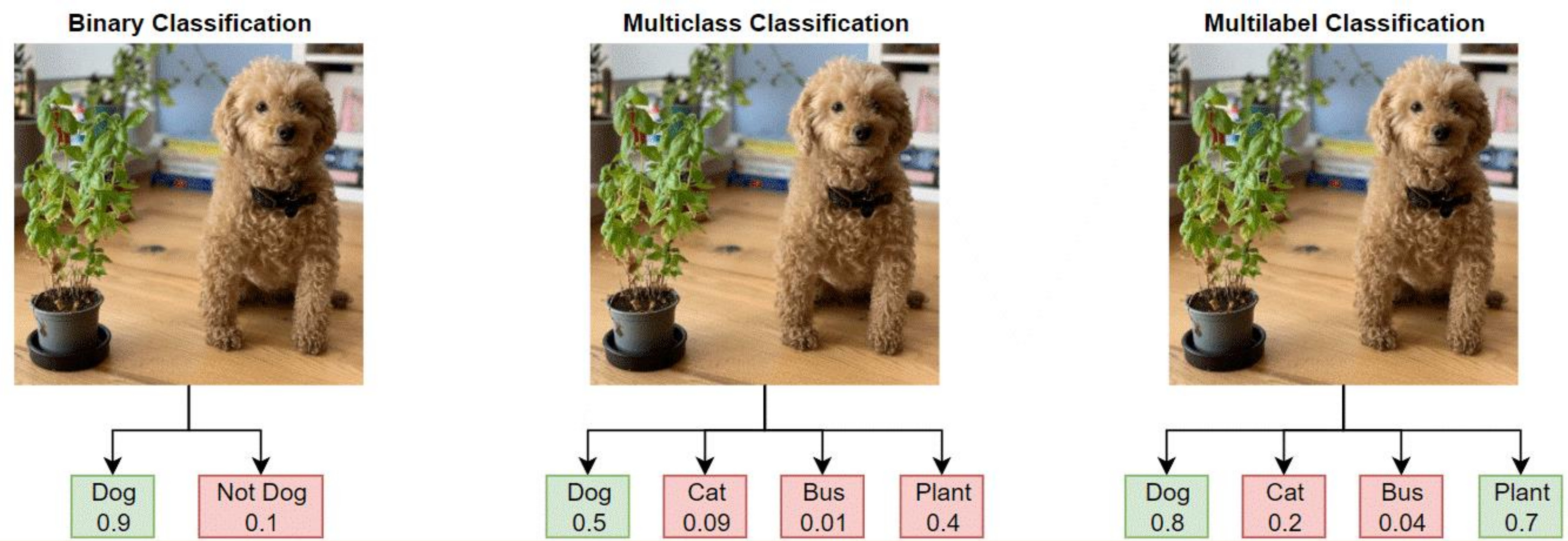
<u>Multi-target:</u>

- Can predict categorical and continous targets as well
- Each instance in the dataset has a fix number of targets
- **Target variables usually have some kind of connection**
- e.g.: predicting the number of customers and revenue; predicting whether customers take an offer and the number of times they use the service etc.

<u>Multi-label:</u>

- Predicts only categorical labels
- Each instance in the dataset can belong to multiple categories, but it is not fixed, how many categories they belong to
- **Target labels usually have some kind of connection**
- e.g.: predicting what objects are on a picture, predicting movie genres, predicting whether one has different diseases etc.

Multi-target vs. multi-label prediction



Different approaches

Problem transformation methods

- transform the problem into separate prediction problems

  one-vs-all OR one-vs-one

- straightforward, but loses information about the relationship of different targets/labels
- e.g.: Binary Relevance, **Classifier Chains**, Conditional Dependency Network, Hierarchy of Multi-label Classifiers (HOMER) etc.

Algorithm adaptation methods

- algorithms are changed in order to take account of label/target correlations
- can be more complex, but does not lose information about target/label relationships
- e.g.: **Predictive Clustering Trees**, different Neural Networks, Twin Multi-Label Support Vector Machine, Multi-label k Nearest Neighbour etc.

Predictive Clustering Trees (PCT)

- Decision tree and random forest modified → TDIDT algorithm
- **Splitting criterion is changed:** Partition data into clusters that are as homogenous as possible in terms of target variables/labels
- Similarity measure: depends on the specific task, for MLC: sum of Gini indicies of the labels
- Prototype and variance function
  - Prototype function: representative summary of a given cluster
  - Variance function: measure of diversity within cluster
- Stopping criteria are the usual suspects: max depth, impurity decrease, max features etc.
- Efficient computation

Measurement of prediction quality

- Output of prediction is more complex than for single-target/binary-label prediction → evaluation is more complex
- But usual scoring methods can be used with some twists

  RMSE, accuracy, precision, F1 score, R^2, Jaccard index, Hamming loss
- Macro- vs. micro-averaging

  per target/label measures are averaged into a single measure vs. the joint value of a statistic is calculated for all targets/labels

  e.g.: MSE with k targets/labels with n observations

<u>Macro:</u> $\sum_j^k w_j \sum_i^n \frac{(y_{ji}-\hat{y}_{ji})^2}{n}$

<u>Micro:</u> $\sum_j^k \sum_i^n \frac{(y_{ji}-\hat{y}_{ji})^2}{kn}$

References

- Jasmin Bogatinovski, Ljupčo Todorovski, Sašo Džeroski, Dragi Kocev (2022): *Comprehensive comparative study of multi-label classification methods, Expert Systems with Applications, Volume 203*
- Dragi Kocev, Celine Vens, Jan Struyf, Sašo Džeroski (2013): *Tree ensembles for predicting structured outputs, Pattern Recognition, Volume 46, Issue 3*
- *https://la.mathworks.com/help/deeplearning/ug/multilabel-image-classification-using-deep-learning.html*