Advanced Machine Learning

Introduction to Multi-Target Prediction

		Tennis	Football	Biking	Movies	TV	Belgium
01101	Text1	0	1	0	0	1	1
00111	Text2	1	0	0	0	0	1
01110	Text3	0	0	0	1	1	0
10001	Text4	0	0	1	0	1	0
01011	Text5	1	0	0	1	0	0
11110	Text6	?	?	?	?	?	?

Learning goals

- Understand the practical relevance of multi-target prediction problems
- Know the relevant special cases of multi-target prediction
- Understand the difference between inductive and transductive learning problems

MULTI-TARGET PREDICTION: MOTIVATION

- Conventional supervised learning: Label/Outcome space $\mathcal{Y}$ is one-dimensional.

$\rightsquigarrow$ The learner predicts *one target variable*, i.e., faces a single-target prediction problem.

- In practice we often have to deal with multi-target prediction (MTP) problems, where multiple target variables of possibly different type (binary, nominal, ordinal, real-valued) need to be predicted.
- Why not just *reducing* an MTP problem to a single-target prediction problem by learning one model per target, independently of the other targets?

plane	person	dog	bus	cat
1	0	0	0	0
0	0	1	0	1
0	1	1	0	0
1	1	0	1	0
1	1	0	0	0
?	?	?	?	?

Iliadis et al. (2021), Multi-target prediction for dummies using two-branch neural networks (URL).

- In most practical cases, the targets are not completely *independent* of each other. Instead, they can be more or less *similar* to each other and exhibit *statistical dependencies*.
- Therefore, one target is (implicitly) also providing *information* about another target, and vice versa.
- Consequently, there is hope to achieve better performance by tackling the targets *simultaneously*.

MULTI-TARGET PREDICTION: CHARACTERISTICS

A multi-target prediction setting is characterized by instances $\mathbf{x} \in \mathcal{X}$ and targets $\mathbf{t} \in \mathcal{T}$ with the following properties:

- **P1** A training data set $\mathcal{D}$ consists of triplets $(\mathbf{x}^{(i)}, \mathbf{t}_j, y_{ij})$, where $y_{ij} \in \mathcal{Y}$ denotes a score that characterizes the relationship between the instance $\mathbf{x}^{(i)} \in \mathcal{X}$ and the target $\mathbf{t}_j \in \mathcal{T}$.
- **P2** In total, n different instances and m different targets are observed during training, with n and m being finite numbers. Thus, the scores y_{ij} of the training data can be arranged in an $n \times m$ matrix Y, which is in general incomplete, i.e., Y may have missing values.
- **P3** The score set $\mathcal{Y}$ is one-dimensional and can be nominal, ordinal or real-valued.
- **P4** The goal is to predict scores for any instance-target pair $(\mathbf{x}, \mathbf{t}) \in \mathcal{X} \times \mathcal{T}$.

In the conventional MTP setting there is no side information for targets available.

Special Cases of Multi-target Prediction

MULTIVARIATE REGRESSION

A multivariate regression problem is a special case of an MTP problem with the following additional properties:

P5 $|\mathcal{T}| = m \rightsquigarrow$ all targets are observed during training.

P6 No side information is available for targets. Without loss of generality, we can hence assign the numbers 1 to m as identifiers to targets, such that the target space is $\mathcal{T} = \{1, ..., m\}$.

P7 The score matrix Y has no missing values.

P8a The score set is $\mathcal{Y} = \mathbb{R}$.

	Mol1	Mol2	Mol3	Mol4	Mol5	Mol6
01101	1,3	0,2	1,4	1,7	3,5	1,3
00111	2	1,7	1,5	7,5	8,2	7,6
01110	0,2	0	0,3	0,4	1,2	2,2
10001	3,1	1,1	1,3	1,1	1,7	5,2
01011	4,7	2,1	2,5	1,5	2,3	8,5
11110	?	?	?	?	?	?

Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

Example: Predict whether a protein (rows) will bind to a set of experimentally developed small molecules (columns).

MULTI-LABEL CLASSIFICATION

A multi-label classification problem is a special case of an MTP problem with the following additional properties:

P5 $|\mathcal{T}| = m \rightsquigarrow$ all targets are observed during training.

P6 No side information is available for targets. Again, without loss of generality, we can hence identify targets with natural numbers, such that the target space is $\mathcal{T} = \{1, ..., m\}$.

P7 The score matrix Y has no missing values.

P8b The score set is $\mathcal{Y} = \{0, 1\}$.

		Tennis	Football	Biking	Movies	TV	Belgium
01101	**Text1**	0	1	0	0	1	1
00111	**Text2**	1	0	0	0	0	1
01110	**Text3**	0	0	0	1	1	0
10001	**Text4**	0	0	1	0	1	0
01011	**Text5**	1	0	0	1	0	0
11110	**Text6**	?	?	?	?	?	?

Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

Example: Assign for documents (rows) the appropriate category tags (columns).

LABEL RANKING

In *label ranking*, each instance is associated with a ranking (total order) of the targets. A label ranking problem is a special case of an MTP problem with the following additional properties:

P5 $|\mathcal{T}| = m \rightsquigarrow$ all targets are observed during training.

P6 No side information is available for targets. Again, without loss of generality, we can hence identify targets with natural numbers, such that the target space is $\mathcal{T} = \{1, ..., m\}$.

P7 The score matrix Y has no missing values.

P8c The score set is $\mathcal{Y} = \{1, \dots, m\}$, and the scores (interpreted as ranks) are such that $y_{ij} \neq y_{ik}$ for all $1 \leq j, k \neq m$.

		Tennis	Football	Biking	Skating	Running	Walking
01101	User 1	2	1	4	3	5	6
00111	User 2	1	4	3	5	6	2
01110	User 3	4	5	1	2	3	6
10001	User 4	4	3	2	6	1	5
01011	User t5	1	3	5	2	6	4
11110	User 6	?	?	?	?	?	?

Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

Example: Predict for users (rows) their preferences over specific activities (columns).

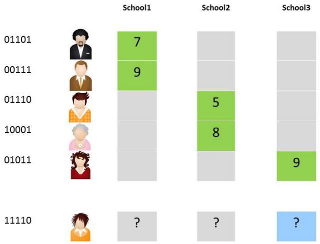
MULTI-TASK LEARNING

A multi-task learning problem is a special case of an MTP problem with the following additional properties:

P5 $|\mathcal{T}| = m \rightsquigarrow$ all targets are observed during training.

P6 No side information is available for targets. Again, the target space can hence be taken as $\mathcal{T} = \{1, ..., m\}$.

P8d The score set is homogenous across columns of Y, e.g., $\mathcal{Y} = \{0, 1\}$ or $\mathcal{Y} = \mathbb{R}$.



Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

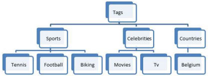
Example: Predict for students (rows) the final grades for a specific high-school course (columns).

Learning with Side Information on Targets

SIDE INFORMATION ON TARGETS

- In some practical applications additional side information about the target space is available.
- Examples:
  - Representation for the target molecules in drug design application (*structured representation*).
  - Taxonomy on document categories (*hierarchy*).

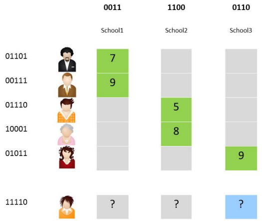
		Mol1	Mol2	Mol3	Mol4	Mol5	Mol6
01101		1,3	0,2	1,4	1,7	3,5	1,3
00111		2	1,7	1,5	7,5	8,2	7,6
01110		0,2	0	0,3	0,4	1,2	2,2
10001		3,1	1,1	1,3	1,1	1,7	5,2
01011		4,7	2,1	2,5	1,5	2,3	8,5
11110		?	?	?	?	?	?



		Tennis	Football	Biking	Movies	Tv	Belgium
01101	Text1	0	0	0	0	0	1
00111	Text2	0	0	1	0	1	1
01110	Text3	0	0	0	1	1	0
10001	Text4	0	0	1	0	1	0
01011	Text5	1	0	0	1	0	0
11110	Text6	?	?	?	?	?	?

SIDE INFORMATION ON TARGETS

- Information about schools and courses (geographical location, qualifications of the teachers, reputation of the school, etc.) in student mark forecasting application (*feature representation*).



Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

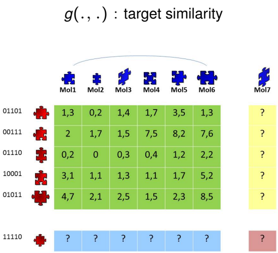
- Such problems are often referred to as dyadic prediction, link prediction, or network inference settings.
- Generally speaking, such settings cover problems that obey the four properties listed in the MTP definition.
- Scores y_{ij} can be arranged in a matrix Y, which is often sparse.
- Thus, *dyadic prediction* can be seen as *multi-target prediction with target features*.

INDUCTIVE VS. TRANSDUCTIVE LEARNING

- In all of the previous problems,
  1. predictions need be be generated for novel instances,
  2. whereas the set of targets is known beforehand and observed during the training phase.

⇝ These problems are *inductive* w.r.t. instances (1) and *transductive* w.r.t. targets (2).

- *Side information* is of crucial importance for generalizing to novel targets that are unobserved during the training phase such as a novel target molecule in the drug design example, a novel tag in the document annotation example, or a novel school in the student grading example.



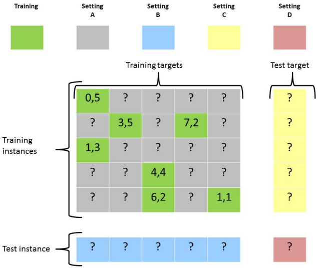
	Mol1	Mol2	Mol3	Mol4	Mol5	Mol6	Mol7
01101	1,3	0,2	1,4	1,7	3,5	1,3	?
00111	2	1,7	1,5	7,5	8,2	7,6	?
01110	0,2	0	0,3	0,4	1,2	2,2	?
10001	3,1	1,1	1,3	1,1	1,7	5,2	?
01011	4,7	2,1	2,5	1,5	2,3	8,5	?
11110	?	?	?	?	?	?	?

Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

SUBDIVISION OF DIFFERENT LEARNING SETTINGS

In light of this, one can distinguish between four different learning settings:

- Setting A — The problem is transductive w.r.t. both the targets and the instances, i.e., both are observed during the training phase, albeit not for all instance-target combinations. The goal is then to predict missing values of the score matrix (*matrix completion problem*).
- Setting B — The problem is transductive w.r.t. the targets and inductive w.r.t. the instances.
- Setting C — The problem is inductive w.r.t. the targets and transductive w.r.t. the instances. This means that some targets are hence not observed during training, but may nevertheless appear at prediction time.
- Setting D — The problem is inductive w.r.t. both the targets and the instances (*zero-shot learning*).



Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).