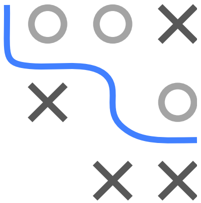
Advanced Machine Learning

Multi-Target Prediction: Introduction

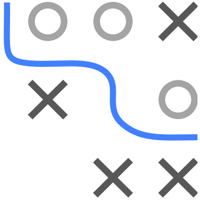


		Tennis	Football	Biking	Movies	TV	Belgium
01101	Text1	0	1	0	0	1	1
00111	Text2	1	0	0	0	0	1
01110	Text3	0	0	0	1	1	0
10001	Text4	0	0	1	0	1	0
01011	Text5	1	0	0	1	0	0
11110	Text6	?	?	?	?	?	?

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

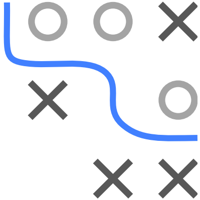


- Conventional supervised learning: Label space $\mathcal{Y}$ is 1-D.
- Multi-target prediction (MTP): multiple targets of mixed types (binary, nominal, ordinal, real-valued).
- Learn one model per target independently? $\rightsquigarrow$ Targets can be *statistically dependent*.
- Multi-label Emotions Dataset: 4 emotions of a music piece. Multiple emotions may be attributed to a single piece. Mutual information of the labels are:

	Calm	**Quiet**	**Sad**	**Angry**
Calm	1.000	0.073	0.018	**0.290**
Quiet	0.073	1.000	**0.241**	**0.164**
Sad	0.018	**0.241**	1.000	0.067
Angry	**0.290**	**0.164**	0.067	1.000

- It might be better to tackle targets *simultaneously*.

MULTI-TARGET PREDICTION: CHARACTERISTICS



Characterized by instances $\mathbf{x} \in \mathcal{X}$ and targets $m \in [1, 2, \dots, l]$ with following properties:

- A training set $\mathcal{D} = \{(\mathbf{x}^{(i)}, y_m^{(i)})\}_{i=1}^n$, where $y_m^{(i)} \in \mathcal{Y}_m$ is label for target m.
- n instances and l targets $\rightsquigarrow$ Scores $y_m^{(i)}$ can be arranged in an $n \times l$ matrix Y. Note Y may have missing values.
- Label space $\mathcal{Y}_m$ can be nominal, ordinal or real-valued.
- Goal: to predict scores for any pair $(\mathbf{x}, m) \in \mathcal{X} \times [1, 2, \dots, l]$.

In conventional MTP setting: no available side information for targets.

Special Cases of Multi-target Prediction

MULTIVARIATE REGRESSION

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

- The score set is $\mathcal{Y} = \mathbb{R}$.
- No side information is available for targets. We can re-index the targets with $\mathbf{t} = (1, 2, \ldots, l)^T$.

	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).

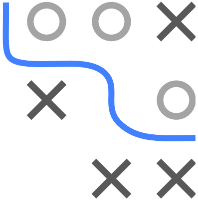
Example: Predict whether a protein (rows) will bind to a set of experimentally developed small molecules (columns).

MULTI-LABEL CLASSIFICATION

- All targets are observed during training.
- No side information is available for targets.
- The score matrix Y has no missing values.
- $\mathcal{Y}_m = \{0, 1\}$ for all m.

		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 of targets.

- All targets are observed during training.
- No side information is available for targets.
- Y has no missing values.
- $\mathcal{Y}_m = \{1, \dots, l\}$ for all m, and the scores (i.e. ranks) $y_m^{(i)} \neq y_k^{(i)}$ for all $1 \leq m, k \neq l$.

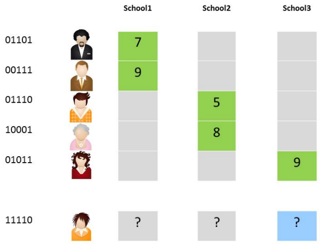
		Tennis	Football	Biking	Skating	Running	Walking
01101	User 1	2	1	4	3	5	6
00111	User 2	1	4	3	5	6	2
01110	User 3	4	5	1	2	3	6
10001	User 4	4	3	2	6	1	5
01011	User t5	1	3	5	2	6	4
11110	User 6	?	?	?	?	?	?

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

- **Not all targets are relavent for all instances**. E.g. a student may only attend one school, other labels are **irrelavent**.
- All targets are observed during training.
- No side information is available for targets.
- Label space is homogenous across columns of Y, e.g., $\mathcal{Y}_m = \{0, 1\}$ or $\mathcal{Y}_m = \mathbb{R}$ for all m.



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).

REMARKS

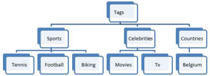
- It is also possible when the m-th task is multiclass classification. That is, $f(\mathbf{x})_m \in \mathbb{R}^{g_m}$ is the probability predictions for g_m classes.
  $\leadsto$ The techniques for multi-target learning are also applicalble under this setting. But the notations will be cumbersome, so we will not cover it.
- Label space can be inhomogeneous. E.g. $\mathcal{Y}_m = \{0, 1\}$ and $\mathcal{Y}_k = \mathbb{R}$.
  $\leadsto$ A mixture of multi-label classification and multivariate regression.

Learning with Side Information on Targets

SIDE INFORMATION ON TARGETS

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

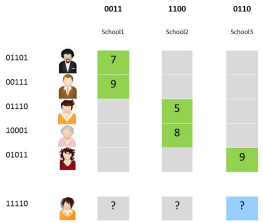
		Mol1	Mol2	Mol3	Mol4	Mol5	Mol6
01101		1,3	0,2	1,4	1,7	3,5	1,3
00111		2	1,7	1,5	7,5	8,2	7,6
01110		0,2	0	0,3	0,4	1,2	2,2
10001		3,1	1,1	1,3	1,1	1,7	5,2
01011		4,7	2,1	2,5	1,5	2,3	8,5
11110		?	?	?	?	?	?



		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

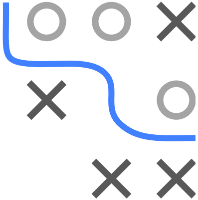
- Information about schools (geographical location, school reputation) in student mark forecasting (*feature representation*).



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

- Such problems are referred to as dyadic or link prediction.
- Such problems fulfill the four properties of MTP setting.
- Scores $y_m^{(i)}$ 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 previous problems,
  1. predictions need be be generated for novel instances,
  2. targets are known beforehand and observed during training.
- These problems are *inductive* w.r.t. instances and *transductive* w.r.t. targets.

- Side information is important for generalizing to novel targets.
  - a novel target molecule in the drug design,
  - a novel tag in the document annotation,

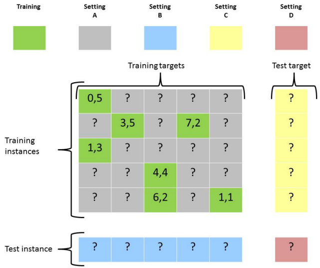
$g(.,.)$: target similarity

	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

- Setting A — transductive w.r.t. targets and instances. Goal: predict missing values of score matrix (*matrix completion*).
- Setting B — transductive w.r.t. targets and inductive w.r.t. instances.
- Setting C — inductive w.r.t. targets and transductive w.r.t. instances. $\rightsquigarrow$ Some targets are unobserved during training but may appear at prediction time.
- Setting D — inductive w.r.t. both targets and instances (*zero-shot learning*).



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