MULTI-TARGET PREDICTION: MOTIVATION

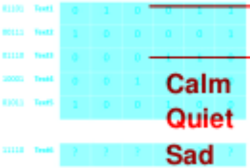
- Conventional supervised learning: Label space $\mathcal{Y}$ is 1-D.
- Multi-target prediction (MTP): multiple targets of mixed types (binary, nominal, ordinal, real-valued).
- Learn one model per target independently? $\rightsquigarrow$ Targets can be *statistically dependent*.
- Multi-label Emotions 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*.

Advanced Machine Learning

Multi-Target Prediction: Introduction



Learning goals

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

MULTI-TARGET PREDICTION: CHARACTERISTICS

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

- A training set $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{n}$, where $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_l^{(i)})$, with $y_m^{(i)} \in \mathcal{Y}_m$ is label for target m.
- n instances and l targets $\rightsquigarrow$ Labels $y_m^{(i)}$ can be arranged in an $n \times l$ matrix $\boldsymbol{Y}$. Note $\boldsymbol{Y}$ may have missing values.
- Target spaces $\mathcal{Y}_m$ can be nominal, ordinal or real-valued.
- Goal: predict scores for any pair $(\mathbf{x}, m) \in \mathcal{X} \times \{1, 2, \ldots, l\}$.

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

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	
Quiet	0.073	1.000		
Sad	0.018		1.000	0.067
Angry			0.067	1.000

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

MULTIVARIATE REGRESSION

Target space $\mathcal{Y}_m = \mathbb{R}\ \forall m \in \{1, 2, \ldots, l\}$.

CHARACTERISTICS

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

- A training set $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{n}$, where $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_l^{(i)})$, with $y_m^{(i)} \in \mathcal{Y}_m$ is label for target m.
- n instances and l targets $\rightsquigarrow$ Labels $y_m^{(i)}$ can be arranged in an $n \times l$ matrix $\boldsymbol{Y}$. Note $\boldsymbol{Y}$ may have missing values.
- Target spaces $\mathcal{Y}_m$ can be nominal, ordinal or real-valued.
- Goal: predict scores for any pair $(\mathbf{x}, m) \in \mathcal{X} \times \{1, 2, \ldots, l\}$.

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

	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 binding strength between proteins (rows) and molecules (columns).

MULTI-LABEL CLASSIFICATION

Target space $\mathcal{Y}_m = \{0, 1\}\ \forall m \in \{1, 2, \ldots, l\}$

		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 documents (rows) to category tags (columns).

LABEL RANKING

In *label ranking*, each instance is associated with a ranking of targets. $\mathcal{Y}_m = \{1, \dots, l\}\ \forall m$, and labels (i.e., ranks) $y_m^{(i)} \neq y_k^{(i)} \forall m \neq k$.

		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 relevent for all instances.** E.g., a student may only attend one school, other labels are **irrelavent**.
- Label space is homogenous across columns of $\boldsymbol{Y}$, 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).

LABEL RANKING

In *label ranking*, each instance is associated with a ranking of targets. $\mathcal{Y}_m = \{1, \dots, l\}\ \forall m$, and labels (i.e., ranks) $y_m^{(i)} \neq y_k^{(i)}\ \forall m \neq k$.

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

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

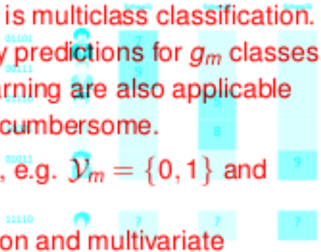
Example: Predict for users (rows) their preferences over specific activities (columns).

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.
  $\rightsquigarrow$ The techniques for multi-target learning are also applicable under this setting, notation becomes cumbersome.
- Target space can be inhomogeneous, e.g. $\mathcal{Y}_m = \{0, 1\}$ and $\mathcal{Y}_k = \mathbb{R}$.
  $\rightsquigarrow$ A mixture of multi-label classification and multivariate regression.

MULTI-TASK LEARNING

- **Not all targets are relevant for all instances.** E.g., a student may only attend one school, other labels are **irrelevant**.
- Label space is homogenous across columns of $\mathbf{Y}$, e.g., $\mathcal{Y}_m = \{0, 1\}$ or $\mathcal{Y}_m = \mathbb{R}$ for all m.

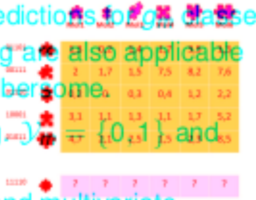


Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

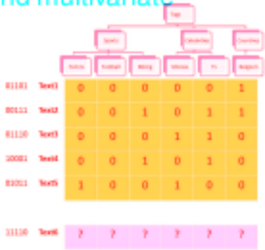
Example: Predict for students (rows) the final grades for a specific high-school course (columns).

SIDE INFORMATION ON TARGETS

- Sometimes, additional side information about targets is available.
- Extra representation for target molecules in drug design (*structured representation*).



- Taxonomy on document categories (*hierarchy*).



That is, $f(\mathbf{x})_m \in \mathbb{R}^{g_m}$ is the probability predictions for g_m classes.
⇝ The techniques for multi-target learning are also applicable under this setting, notation becomes cumbersome.

- Target space can be inhomogeneous, e.g. $\mathcal{Y}_1 = \{0, 1\}$ and $\mathcal{Y}_k = \mathbb{R}$.
⇝ A mixture of multi-label classification and multivariate regression.

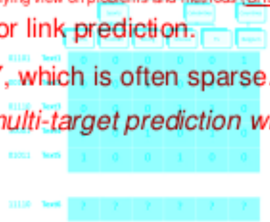
SIDE INFORMATION ON TARGETS / 2

- Sometimes, additional side information about targets is available.

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

- Extra representation for target molecules in drug design (*structured representation*).

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

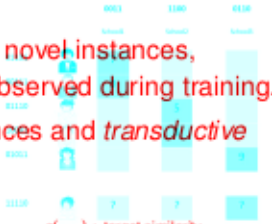


- Such problems are referred to as dyadic or link prediction.
- Labels $y_m^{(i)}$ can be arranged in a matrix $\boldsymbol{Y}$, which is often sparse.
- Thus, *dyadic prediction* can be seen as *multi-target prediction with target features.*

- Taxonomy on document categories (*hierarchy*).

INDUCTIVE VS. TRANSDUCTIVE LEARNING

- In previous problems,
  1. predictions need to 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.



$g(.,.)$: target similarity

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

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

- Information about schools (geographical location, school reputation) in student mark forecasting (*feature representation*).
- Such problems are referred to as dyadic or link prediction.
- Labels $y^{(i)}$ can be arranged in a matrix Y, which is often sparse.
- Thus, *dyadic prediction* can be seen as *multi-target prediction with target features*.

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 (*classical supervised learning*).
- Setting C — inductive w.r.t. targets and transductive w.r.t. instances.
  ⇝ 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*).

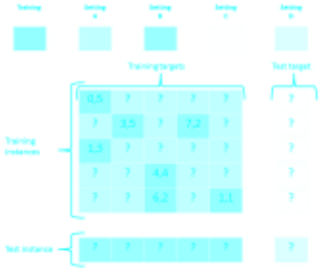
Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods (URL).

- In previous problems,
  - predictions need to be generated for novel instances;
  - 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,

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 (*classical supervised learning*).
- Setting C — inductive w.r.t. targets and transductive w.r.t. instances. ⇝ 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).