

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
01100	Test1	0 1 0 0 0 0 0				
00111	Test2	1 0 0 0 0 0 1				
01110	Test3	0 0 0 0 0 0 0				
10000	Test4	0 0 0 1 0 0 0	Calm	1.000	0.073	0.018
01001	Test5	1 0 0 0 0 0 0	Quiet	0.073	1.000	0.241
01110	Test6	0 0 0 0 0 0 0	Sad	0.018	0.241	1.000
			Angry	0.290	0.164	0.067

- 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:

- Multi-target prediction (MTP): multiple targets of mixed types
- A training set $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^n$, where $\mathbf{y}^{(i)} = (y_1^{(i)}, \dots, y_l^{(i)})$, with (binary, nominal, ordinal, real-valued).
- $y_m^{(i)} \in \mathcal{Y}_m$ is label for target m .
- Learn one model per target independently? \rightsquigarrow Targets can be statistically dependent.
- n instances and l targets \rightsquigarrow Labels $y_m^{(i)}$ can be arranged in an $n \times l$ matrix \mathbf{Y} . Note \mathbf{Y} may have missing values.
- Multi-label Emotions Dataset: 4 emotions of a music piece.
- Target spaces \mathcal{Y}_m can be nominal, ordinal or real-valued.
- Multiple emotions may be attributed to a single piece. Mutual information of the labels are:
- Goal: 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.

	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 CHARACTERISTICS

Target space $\mathcal{Y}_m \subseteq \mathbb{R} \cup \{1, 2, \dots, l\}$ for targets $m \in \{1, 2, \dots, l\}$ with following properties:

- A training set $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^n$, where $y^{(i)} = (y_1^{(i)}, y_2^{(i)}, \dots, y_l^{(i)})$, 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 Y . Note Y may have missing values.
- Target spaces \mathcal{Y}_m can be nominal, ordinal or real-valued.
- Goal: predict scores for any pair $(x, m) \in \mathbb{R}^d \times \{1, 2, \dots, l\}$.

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

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	2,5,2
01011	+	1,7	2,1	1,5	1,5	1,3	3,5
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 $\mathcal{Y}_m = \{0, 1\} \forall m \in \{1, 2, \dots, l\}$



		Tennis	Football	Biking	Movies	MMS	Religion
01101	Text1	103	012	104	107	315	113
00111	Text2	1	107	105	705	802	716
01110	Text3	002	0	003	014	112	202
10001	Text4	301	101	113	101	117	502
01011	Text5	417	201	205	115	203	805
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).
molecules (columns).

LABEL RANKING CLASSIFICATION

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	Movies	TV	Belgium
		Tennis	Football	Biking	Skating	Running	Walking
01101	User 1	2	0	1	4	3	5
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 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).



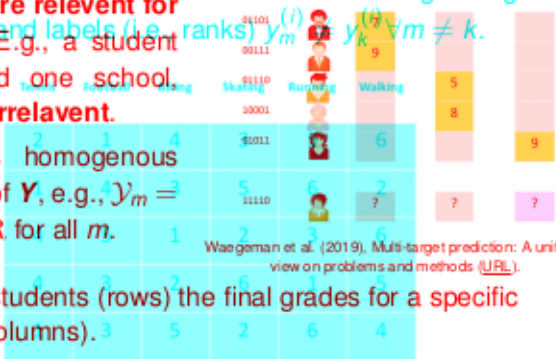
MULTI-TASK LEARNING

In *label ranking*, each instance is associated with a ranking of targets.

- Not all targets are relevant for all instances. E.g., a student may only attend one school, other labels are irrelevant.

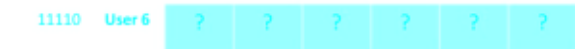
$$\mathcal{Y}_m = \{1, \dots, k\} \cup \emptyset, \text{ and labels (i.e., ranks) } y_m^{(i)} \neq y_k^{(i)} \text{ if } \forall m \neq k.$$

- 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 .



Waegele 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).



Waegele 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 ON LEARNING

- 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.
 - 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 $\mathcal{Y}_k = \mathbb{R}$.
 - A mixture of multi-label classification and multivariate regression.

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. That is, $f(x)_m \in \mathbb{R}^{g_m}$ is the probability predictions for g_m classes.
 - The techniques for multi-target learning are also applicable under this setting, notation becomes cumbersome.
- Extra representation for target molecules in drug design (*structured representation*).
- Target representation inhomogeneous, e.g. $\mathcal{Y}_k = \mathbb{R}$.
 - A mixture of multi-label classification and multivariate regression.
- Taxonomy on document categories (*hierarchy*).



	MAA1	MAA2	MAA3	MAA4	MAA5	MAA6
10001	2	1.7	3.5	7.5	8.2	7.6
10002	0.2	0.4	3.2	2.2		
10003	3.1	1.1	3.3	1.1	2.7	5.2
10004	4.7	1.1	2.5	2.5	3.1	8.5



SIDE INFORMATION ON TARGETS / 2

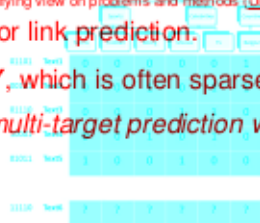
- Sometimes, additional side information about targets is available.

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



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 $y_{ij}^{(i)}$ can be arranged in a matrix \mathbf{Y} , which is often sparse.
- This dyadic prediction can be seen as multi-target prediction with target features.

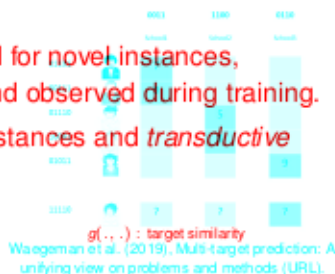


INDUCTIVE VS. TRANSDUCTIVE LEARNING

- In previous problems,
 - predictions need to be generated for novel instances,
 - information about schools (geographical location, school reputation) is 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.
- Such problems are referred to as *dyadic* or *link prediction*.
- Labels $y^{(i)}$ can be expressed in a matrix Y , which is often sparse.
- Thus, *dyadic prediction* can be seen as *multi-target prediction with target features*.
 - a novel target molecule in the drug design,
 - a novel tag in the document annotation,



SUBDIVISION OF DIFFERENT LEARNING SETTINGS

- In previous problems,
- Setting A — transductive w.r.t. targets and instances. Goal: predict missing values of score matrix (*matrix completion*).
 - predictions need to be generated for novel instances, targets are known beforehand and observed during training.
- Setting B — transductive w.r.t. targets and inductive w.r.t. instances (*classical supervised learning*).
 - these problems are inductive w.r.t. instances and transductive w.r.t. targets.
- Setting C — inductive w.r.t. targets and transductive w.r.t. instances.
 - side information is important for generalizing to novel targets.
 - Some targets are unobserved during training but may appear at prediction time.
 - a novel target molecule in the drug design.
- Setting D — inductive w.r.t. both targets and instances (*zero-shot learning*).
 - a novel tag in the document annotation.

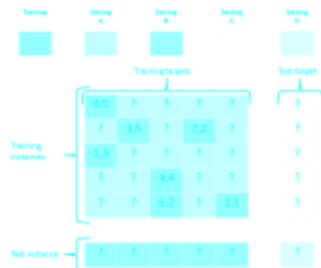


Waegeman et al. (2019), Multi-target prediction: A unifying view on problems and methods ([URL](#)).

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](#)).

