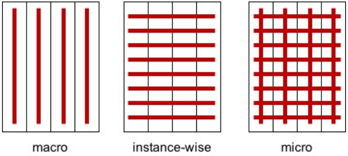
Advanced Machine Learning

Loss Functions for Multi-Target Prediction



Learning goals

- Get to know loss functions for multi-target prediction problems
- Know the Bayes predictor for Hamming loss and subset $0/1$ loss
- Understand the difference between macro-, micro-, and instance-wise-losses

MULTIVARIATE LOSS FUNCTIONS

- In multi-target prediction we want to the following: For a feature vector **x**, predict a vector of scores $\mathbf{y} = (y_1, y_2, \ldots, y_m)^\top$ by means of a function (hypothesis) f:

$$\mathbf{x} = (x_1, x_2, \ldots, x_p)^\top \quad \xrightarrow{f(\mathbf{x})} \quad \hat{y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m)^\top$$

- If we want to follow the machine learning paradigm based on loss minimization, we need a *multivariate loss functions*

$$\ell : \mathcal{Y}^m \times \mathcal{Y}^m \to \mathbb{R}.$$

  Compared to single-target prediction, a broad spectrum of such multivariate loss functions is conceivable.

- In case we have an appropriate multivariate loss function ℓ, we want to find a (Bayes) predictor f^* that minimizes expected loss with regard to ℓ :

$$\begin{aligned} f^* &= \arg\min_{f:\mathcal{X}\to\mathcal{Y}^m} \mathcal{R}_\ell(f) = \arg\min_{f:\mathcal{X}\to\mathcal{Y}^m} \mathbb{E}_{xy}\left[\ell(y, f(\mathbf{x}))\right] \\ &= \arg\min_{f:\mathcal{X}\to\mathcal{Y}^m} \int \ell(y, f(\mathbf{x}))\mathrm{d}\mathbb{P}_{xy}. \end{aligned}$$

EXAMPLES OF MTP LOSS FUNCTIONS

- *Squared error loss* (typically used in multivariate regression):

$$\ell(\mathbf{y}, \hat{y}) = \sum_{j=1}^{m} (y_j - \hat{y}_j)^2 \,,$$

  where $\mathbf{y}, \hat{y} \in \mathbb{R}^m$.

- The *Hamming loss* averages over mistakes on individual scores:

$$\ell_H(\mathbf{y}, \hat{y}) = \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}_{[y_j \neq \hat{y}_j]}$$

- The *subset 0/1 loss* simply checks for entire correctness:

$$\ell_{0/1}(\mathbf{y}, \hat{y}) = \mathbb{1}_{[\mathbf{y} \neq \hat{y}]} = \max_j \mathbb{1}_{[y_j \neq \hat{y}_j]}$$

HAMMING VS. SUBSET 0/1 LOSS

- The risk minimizer for the Hamming loss is the *marginal mode*:

$$f_j^*(\mathbf{x}) = \arg \max_{y_j \in \{0,1\}} \Pr(y_j \mid \mathbf{x}), \quad j = 1, \ldots, m,$$

while for the subset 0/1 loss it is the *joint mode*:

$$\mathbf{f}^*(\mathbf{x}) = \arg \max_{\mathbf{y} \in \mathcal{Y}^m} \Pr(\mathbf{y} \mid \mathbf{x}).$$

- Marginal mode vs. joint mode:

$\mathbf{y}$	$\Pr(\mathbf{y})$
0 0 0 0	0.30
0 1 1 1	0.17
1 0 1 1	0.18
1 1 0 1	0.17
1 1 1 0	0.18

Marginal mode: 1 1 1 1
Joint mode: 0 0 0 0

MULTIVARIATE LOSS FUNCTIONS

- A loss L (on test data) is decomposable over examples if it can be written in the form

$$L = \sum_{i=1}^{n} \ell(\mathbf{y}^{(i)}, f(\mathbf{x}^{(i)}))\,,$$

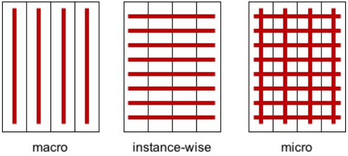
  i.e., as a sum of losses over all (test) examples.

- A multivariate loss ℓ is decomposable over targets if it can be written as

$$\ell(\mathbf{y}, f(\mathbf{x})) = \sum_{j=1}^{m} \ell_j(y_j, f_j(\mathbf{x}))$$

  with suitable single-target losses ℓ_j.

- In general, we distinguish between three categories of losses: macro-, micro-, and instance-wise-losses.



MACRO- AND MICRO-LOSSES

- Macro-losses: The overall loss corresponds to aggregating the losses over the targets.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging the target losses.

$$L = \frac{1}{4}\left(L_1 + L_2 + L_3 + L_4\right)$$

MACRO- AND MICRO-LOSSES

- Macro-losses: The overall loss corresponds to aggregating the losses over the targets.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging the target losses.

$$L = \frac{1}{4}\left(L_1 + L_2 + L_3 + L_4\right)$$

MACRO- AND MICRO-LOSSES

- Macro-losses: The overall loss corresponds to aggregating the losses over the targets.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging the target losses.

$$L = \frac{1}{4}(L_1 + L_2 + L_3 + L_4)$$

MACRO- AND MICRO-LOSSES

- Macro-losses: The overall loss corresponds to aggregating the losses over the targets.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging the target losses.

$$L = \frac{1}{4}\left(L_1 + L_2 + L_3 + L_4\right)$$

MACRO- AND MICRO-LOSSES

- Macro-losses: The overall loss corresponds to aggregating the losses over the targets.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging the target losses.

$$L = \frac{1}{4}(L_1 + L_2 + L_3 + L_4)$$

MACRO- AND MICRO-LOSSES

- Micro-losses: The overall loss corresponds to aggregating the pointwise losses over the targets and the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Thus, we have

$$L = \sum_{i,j} \ell(y_{ij}, \hat{y}_{ij}),$$

where $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ in this case.

MACRO- AND MICRO-LOSSES

- Micro-losses: The overall loss corresponds to averaging the pointwise losses over the targets and the instances.

True scores

y_{11}	y_{12}		y_{14}
y_{21}		y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}		y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
	y_{62}	y_{63}	

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$		$\hat{y}_{14}$
$\hat{y}_{21}$		$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$		$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
	$\hat{y}_{62}$	$\hat{y}_{63}$	

- Thus, we have

$$L = \sum_{i,j} \ell(y_{ij}, \hat{y}_{ij}),$$

where $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ in this case.

- Can be used also for cases with missing entries.

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \\ \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$

INSTANCE-WISE LOSSES

- Instance-wise losses: Aggregating the losses over the instances.

True scores

y_{11}	y_{12}	y_{13}	y_{14}
y_{21}	y_{22}	y_{23}	y_{24}
y_{31}	y_{32}	y_{33}	y_{34}
y_{41}	y_{42}	y_{43}	y_{44}
y_{51}	y_{52}	y_{53}	y_{54}
y_{61}	y_{62}	y_{63}	y_{64}

Predicted scores

$\hat{y}_{11}$	$\hat{y}_{12}$	$\hat{y}_{13}$	$\hat{y}_{14}$
$\hat{y}_{21}$	$\hat{y}_{22}$	$\hat{y}_{23}$	$\hat{y}_{24}$
$\hat{y}_{31}$	$\hat{y}_{32}$	$\hat{y}_{33}$	$\hat{y}_{34}$
$\hat{y}_{41}$	$\hat{y}_{42}$	$\hat{y}_{43}$	$\hat{y}_{44}$
$\hat{y}_{51}$	$\hat{y}_{52}$	$\hat{y}_{53}$	$\hat{y}_{54}$
$\hat{y}_{61}$	$\hat{y}_{62}$	$\hat{y}_{63}$	$\hat{y}_{64}$

- Example: Averaging over the instance-losses.

$$L = \frac{1}{6}\Big(\ell(\mathbf{y}^{(1)}, \hat{y}^{(1)}) + \ell(\mathbf{y}^{(2)}, \hat{y}^{(2)}) + \ell(\mathbf{y}^{(3)}, \hat{y}^{(3)}) + \ell(\mathbf{y}^{(4)}, \hat{y}^{(4)}) + \ell(\mathbf{y}^{(5)}, \hat{y}^{(5)}) + \ell(\mathbf{y}^{(6)}, \hat{y}^{(6)})\Big)$$