Submissions will be evaluated by SWC using the multi-class logarithmic loss and each sample has been labelled with one true class. For each sample, you must submit a set of predicted probabilities (one for every class). The formula is then

$$logloss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij}),$$

where **N** is the number of samples in the test set, **M** is the number of class labels, **log** is the natural logarithm, $\mathbf{y_{ij}}$ is 1 if observation i is in class j and 0 otherwise, and $\mathbf{p_{ij}}$ is the predicted probability that observation i belongs to class j. The submitted probabilities for a given sample are not required to (but are encouraged to) sum to one because they will be re-scaled prior to being scored (each row is divided by the row sum), but they need to be in the range of [0, 1]. To avoid the extremes of the log function, predicted probabilities are replaced with

$$max(min(p, 1 - 10^{-15}), 10^{-15})$$

Log-loss increases as the predicted probability diverge from the actual label and decreases as the predicted probability gets closer to the actual label. The goal of our model is to minimize log loss value. A perfect model would have a log loss of 0. The aim here is to keep the value as low as possible. If the model divides an observation into a wrong class and is very sure that it is correct, it gets a higher penalty than if it divides into a wrong class, but with a not so high conviction (percentage value). The same is true for correct classification. The lowest loss is achieved by a classification into the correct class with 100% certainty, the highest loss is achieved by a classification into a wrong class with a 100% certainty. Log loss penalizes both types of errors, but especially those predictions that are **confident and wrong**!