coursera

Question 1

Suppose we solve a binary classification task and our solution is scores with logloss. What predictions are more preferable in terms of logloss if true labels are y_true = [0, 0, 0, 0].

Correct answers:

- y_pred = [0.5, 0.5, 0.5, 0.5].

Incorrect answers:

- y_pred = [0, 0, 0, 1]. Incorrect! Technically, the loss is infinite in this case, while is it is not for other options, so it cannot be the right answer.
- y_pred = [0.4, 0.5, 0.5, 0.6]. Incorrect! What is better to predict [0.5, 0.5] or [0.4, 0.6]? To answer this question think how $\log$ function behaves. If you plot it you will clearly see that $\log(6) - \log(5) < log(5) - log(4)$, thus $log(4) + log(6) < log(5) + log(5)$. In fact it follows from Jensen's inequality.

Question 2

Suppose we solve a regression task and we optimize MSE error. If we managed to lower down MSE loss on either train set or test set, how did we change Pearson correlation coefficient between target vector and the predictions on the same set?

Correct answers:

- Any behavior is possible. Yes! We cannot monotonically relate MSE and Pearson correlation similarly to how e.g. R-squared monotonically related MSE,

Incorrect answers:

- The correlation did not change. Try to come up with a counterexample when correlation is zero, but MSE can be lowered down.