

⚠ Try again once you are ready

Grade received 43.05% To pass 80% or higher

Try again

Continuous Evaluation and Monitoring

Total points 6

1. Many ML models suffer from declining predicting capabilities over time. A common solution used to overcome this deterioration is to keep retraining your model with new data. As part of this process you may encounter a phenomenon called concept emergence. Which of the following statements accurately describes this emergent phenomenon?

1 / 1 point

- ☒ The appearance of new patterns in data distribution that were not previously present in your dataset.
- ☐ The loss of prediction quality over time.
- ☐ The persistent appearance of stationary data that remains immutable over time.
- ☐ The lack of covariate shift.

✔ Correct

That's right! These new patterns are the new concepts emerging in your data. Continuous evaluation and monitoring will help you address them properly.

2. Statistical process control is a technique that detects concept drift assuming that the errors follow a binomial distribution. Would the system trigger an alarm if $p_t = \sigma_t = 0.3$ and $p_{min} = \sigma_{min} = 0.12$?

0 / 1 point

- ☐ Yes
- ☒ No

✘ Incorrect

Not quite! Please review your calculations and/or the alert rule for this concept drift test.

3. In sequential analysis you detect concept drift by calculating the negative predictive value, precision, recall, and specificity of the system based on a standard contingency table. If the data is stationary these quantities should not change over time. This analysis is tedious as it requires recomputing all these metrics each time we get a new sample. Which of the following approaches is usually implemented to overcome this problem?

0 / 1 point

- ☐ Monte Carlo sampling
- ☒ Adaptive windowing
- ☐ Recursive computation and caching
- ☐ Incremental update rule

✘ Incorrect

Not quite! This still requires computing multiple instances based on multiple contingency tables.

4. Drift detection techniques in unsupervised settings typically suffer from the curse of dimensionality. Which of the following techniques is an appropriate solution to mitigate the effects of this curse? (Check all that apply)

0.75 / 1 point

- ☒ SVD (Singular Value Decomposition)

✘ This should not be selected

Not quite! SVD is a powerful matrix factorization but you need to perform an additional step based on the SVD to avoid the dimensionality curse.

- ☒ PCA (Principal components analysis)

✔ Correct

That's right! Principal components analysis is the right tool to reduce the number of features to detect drift more efficiently.

- ☒ NMF (Non Negative Matrix Factorization)

✔ Correct

That's right! NMF is a very useful dimensionality reduction technique when the features are constrained to be all non-negative.

- ☐ K-means

5. In unsupervised settings, clustering is a very useful method to detect novelty in your data. In this method, you cluster the incoming batches of data to one of the known classes. If you observe that the features of the new data are lying far away from the classes of known features, you can term it as an emerging concept. The downside of this method is that it detects only _____ drift and not _____ changes.

0 / 1 point

- ☐ cluster-based, population-based.
- ☒ feature, cluster-based
- ☐ cluster-based, feature
- ☐ population-based, cluster-based

✗ Incorrect

Not quite! The method assigns new data to known classes so that's the main limitation.

6. It is a sad truth that most of the machine learning models are trained with a fixed set of stationary data. It is very likely that in this process you may have slightly biased your model in favor of your limited data at training. Consequently, as time progresses, your ML model's performance will deteriorate with time. Monitoring helps prevent this performance decay in which ways? (Check all that apply)

0.8333333333333334 / 1 point

- ☒ By retraining your model constantly

✗ This should not be selected

Not quite! Ideally you want to set up mechanisms to detect drift automatically and retrain based on this assessment.

- ☒ Allows you to identify distribution changes close to the classification boundaries

✓ Correct

Yes! Data drift can change classification boundaries quite drastically and monitoring will help you detect and mitigate this unwanted behavior.

- ☐ By performing dimensionality reduction

- ☐ Allows you to establish ground truth labels

- ☒ Identify regions in latent space where the model performs poorly

✓ Correct

That's right! Monitoring will help you identify areas in latent space where your model struggles at classification. You can further use this knowledge to refine your model.

- ☒ Reduces false alarm rates

✓ Correct

Correct! If applied wisely, monitoring can detect data drift early on and adjust the model accordingly to adapt to these changes and hence improving model's performance.