ML Review Quiz: Big Recap

* Required

1. 3 main types of learning * 3 points

You were introduced to 3 different learning strategies utilized by machine learning algorithms: unsupervised learning, supervised learning, and reinforcement learning. Match each type of learning to one of the goals below:

Mark only one oval per row.

	unsupervised	supervised	reinforcement
Learn how to optimally behave in your environment	◯	◯	◯
Discover patterns in your data	◯	◯	◯
Predict Y from X	◯	◯	◯

2. The main ML types also have subcategories... * 5 points

...match each subcategory to its main learning type.

Mark only one oval per row.

	Unsupervised Learning	Supervised Learning
Dimensionality Reduction	◯	◯
Clustering	◯	◯
Anomaly Detection	◯	◯
Regression	◯	◯
Classification	◯	◯

3. Assign each algorithm to one of these categories * 11 points

Mark only one oval per row.

	Dimensionality Reduction	Anomaly Detection	Clustering	Features-based Supervised Learning Models	Similarity-based Supervised Learning Models
PCA					
t-SNE					
k-means					
DBSCAN					
gamma-index					
linear / logistic regression					
Decision Tree					
Random Forest					
Neural Network					
k-nearest-neighbors					
SVM (a kernel method)					

4. Data vs. Concept Drift *

2 points

In this task we want to predict the price of a house from its size. Which of the scenarios results in either a data or a concept drift? Remember: Data Drift is when the input distribution changes, Concept Drift is when the input/output relation changes.

Mark only one oval per row.

	Data Drift	Concept Drift
Due to inflation, the same size house now costs 20% more than 5 years ago.	◯	◯
On average, people built larger houses in 2020 than in 2010.	◯	◯

5. Over- and Underfitting *

2 points

A model's performance on new data points can be bad for two reasons: underfitting (= the model has a high bias) or overfitting (= high variance). Which is which?

Mark only one oval per row.

	underfitting	overfitting
When you evaluate the model on the data it was trained on, the performance is close to that of a human, but on new data points it performs poorly.	◯	◯
No matter on what data (train or test) you evaluate the model, the performance is always far below that of a human.	◯	◯

6. Improving the performance * 5 points

Which of these actions can help if your model is either over- or underfitting?

Mark only one oval per row.

	underfitting	overfitting
try a more complex (non-linear) model	◯	◯
feature engineering	◯	◯
use regularization	◯	◯
get more data (samples)	◯	◯
feature selection	◯	◯

7. Machine learning is an “iterative” process, meaning that an AI team often has to try many ideas before arriving at a solution that’s good enough, rather than have the first thing they try work. * 1 point

Mark only one oval.

- ◯ True
- ◯ False

8. Which of these are reasons that it’s often unrealistic to expect an ML system to be 100% accurate? * 1 point

Mark only one oval.

- ◯ You might not have enough data
- ◯ Data can be mislabeled
- ◯ Data can be ambiguous
- ◯ All of the above