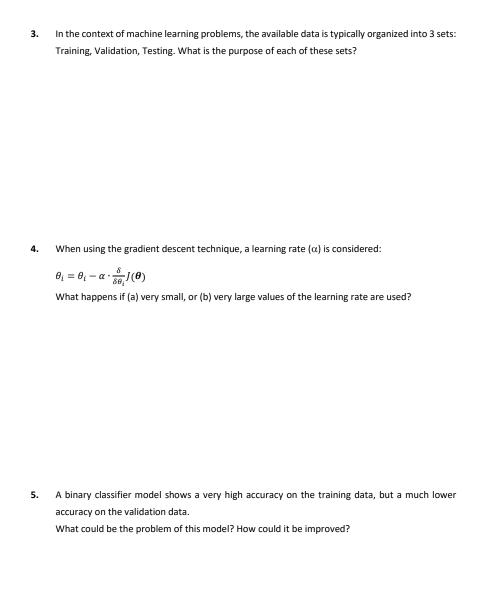


Learning-based Multimedia Processing

2021/2022

Quiz #4	
IST number:	
Name:	
	Duration: 20 minutes Provide clear, legible, and succinct answers. Always justify your assumptions.
	Questions
1.	The two main classes of supervised machine learning algorithms are regression and classification. Give one application example, and a possible technique to use, for each of these classes.

2. Accuracy, defined as $\frac{TP+TN}{TP+TN+FP+FN'}$ is a performance metric that is often used. Is it appropriate for usage with unbalanced datasets? Why?



```
Solutions
```

1

Regression: estimating house prices – linear regression;

Classification: face verification – SVM.

2

No. When the number of positive samples (P=TP+FN) is very small then the accuracy becomes $\frac{TN}{TN+FP} = \frac{TN}{N}$, not reflecting the system performance (balanced accuracy could be used instead - Balanced accuracy = $\frac{TPR+TNR}{2}$, where $TPR = \frac{TP}{P}$ and $TNR = \frac{TN}{N}$)

3

Training dataset – examples used to fit the model parameters.

Validation dataset – used to choose the model's hyperparameters.

Test dataset – used to provide an unbiased evaluation of a final model.

4

- (a) small learning rate: slow convergence;
- (b) large learning rate: may not convergence

5

The model is overfitting the data – Use more training data; use lower order/simpler model; use regularization