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?

3. In the context of machine learning problems, the available data is typically organized into 3 sets: Training, Validation, Testing. What is the purpose of each of these sets?

4. When using the gradient descent technique, a learning rate (α) is considered:

$$\theta_i = \theta_i - \alpha \cdot \frac{\delta}{\delta\theta_i} J(\boldsymbol{\theta})$$

What happens if (a) very small, or (b) very large values of the learning rate are used?

5. A binary classifier model shows a very high accuracy on the training data, but a much lower accuracy on the validation data.
What could be the problem of this model? How could it be improved?

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