Assignment 6 (Knowledge Discovery and Data Mining)

Number of problems/points: Four problems for total of 100 points
Out: February 27, 2020
Due: March 12, 2020 in class (hard copy).

**Problem 1: (25 points)**
Solve Census Income problem from Homework 5 using SVM model and report F1 score.

**Problem 2: (25 points)**
Solve Census Income problem from Homework 5 using Random Forest model and report F1 score.

**Problem 3: (20 points)**
(a) Illustrate on an example the vanishing gradient problem for a deep neural network (with many hidden layers) if using sigmoid activation function.
(b) What is a way to overcome this problem (explain how)?

**Problem 4: (30 points)**
Propose three topics on which you would like to present in class and for each topic provide a reference to materials which you will use to prepare the presentation. These topics should be different from topics already discussed in class. Each topic should be appropriate for a 15 minutes presentation. You can present based on materials from two textbooks, but you are also allowed to use other books, conference tutorial slides, articles etc). Following are possible topics to consider (you can also consider different topics):

- Large scale hierarchical classification
- Advanced concepts in cluster analysis
- Association rules mining
- Advanced concepts in association analysis
- Anomaly detection
- Data stream mining
- Text and web mining
- Time series mining
- Mining big time series
- Sequence pattern mining
- Survival analysis
- Mining spatial data
- Mining graphs
- Graphs sketching, sampling, streaming
- Mining web data
- Mining social networks
- Privacy preserving data mining
- Mining spatio-temporal data
- Mining semistructured data
- Mining with constraints
- False discoveries
- Lifelong machine learning
- Deep Bayesian mining
- Data mining for drug discovery
- Mining electronic health records
• Data mining in transportation
• Data mining in power systems
• Sports analytics
• Explainable data modeling
• Active learning
• Human-in-the-loop learning
• Visual analytics
• Fairness-aware machine learning
• Transfer learning
• Fake news detection
• Zero-shoot learning
• Mining temporal networks
• Reinforcement learning
• Graph neural networks
• Deep reinforcement learning
• Deep learning for personalized search and recommender systems
• A/B testing at scale