**THEORETICAL DATA SCIENCE SYLLABUS**

**Regression algorithms:**

1. The regression line
2. Fitting polynomials
3. Overfitting and regularization
4. The Kernel trick
5. One-hot-encoding

**Optimization and the Gradient Descent method:**

1. Linear regression by gradient descent
2. Early stopping of gradient descent
3. Gradient descent with momentum

**Classification algorithms:**

1. Logistic regression, the sigmoid function, and the binary cross-entropy loss function
2. Softmax regression and the categorical cross-entropy loss function
3. K-nearest neighbors (KNN) method

**Clustering algorithms:**

1. K-means
2. Hierarchical clustering
3. Spectral clustering
4. Gaussian mixtures and the maximization-expectation method

**Recommender systems:**

1. Item-based recommendations. The Jaccard, serendipity, cosine and correlation similarities.
2. User-based recommendations
3. Matrix factorization models

**Neural Networks:**

1. Dense and activation layers.
2. Loss (mean square error and categorical cross entropy) layers
3. Backpropagation
4. The ADAM (adaptive momentum) optimizer
5. Dropout layers
6. Convolutional Neural Networks (CNNs)
7. Recurrent Neural Networks.

**Anomaly detection methods**

**Dimensionality reduction methods**

1. The curse of dimensionality
2. Principal components analysis
3. Locally linear embedding
4. T-Stochastic Neighbor Embedding (t-SNE)
5. Uniform manifold approximation and projection (UMAP)

**Graphs and networks**