Summary of Lecture 1: Deep Learning with Applications to Finance

January 21, 2025

Key Topics Covered

Introduction to R

R is introduced as a general-purpose statistical programming language widely used in both academia and industry. Key features include:

- Handling numerical, character, and logical data types.
- Vectorized operations and summary statistics such as `mean`, `var`, and `sd`.
- Data structures like vectors, lists, matrices, and data frames for organizing and manipulating data.
- Built-in functions for visualization, including histograms and scatter plots.

Data Aggregation and Visualization

The course emphasizes:

- Aggregating data with functions like `aggregate`.
- Analyzing built-in datasets, such as the Iris dataset, using tools for statistical summaries and visual representation.
- Sorting and filtering data using logical operators and indexing.

Introduction to Deep Learning

Deep learning is presented as a subset of machine learning, leveraging neural networks to model complex patterns in data. Highlights include:

- **Supervised Learning**: Techniques for regression, binary classification, and categorical classification.
- **Activation Functions**: Linear, sigmoid, and softmax functions for various prediction tasks.
- **Loss Functions**: Mean squared error (MSE) for regression and cross-entropy for classification tasks.

Overfitting and Regularization

Strategies to address overfitting include:

- Early stopping based on validation performance.
- Regularization techniques such as Ridge (L2), Lasso (L1), and Elastic Net penalties.
- Dropout layers to randomly omit input signals during training.

Optimization Techniques

The course discusses optimization methods central to training deep learning models:

- **Gradient Descent**: Standard algorithm for updating weights.
- **Advanced Optimizers**: RMSProp, Adam, and stochastic weight averaging (SWA).
- Techniques like learning rate decay and momentum to improve convergence.

Pipeline for Deep Learning

Steps for implementing a deep learning project are outlined:

- Data preparation, including normalization and handling missing data.
- Model selection, hyperparameter tuning, and validation using techniques like cross-validation.
- Deployment considerations, including meeting business goals for speed and accuracy.

Practical Applications

Students work on hands-on tasks such as:

- Standardizing vectors and calculating variance without built-in functions.
- Using the Iris dataset to compute species-wise statistics and generate data frames for specific criteria.
- Implementing Keras models for supervised learning tasks in R.

Conclusion

The course provides a robust foundation in statistical programming and deep learning, with a focus on practical applications in finance. Students gain experience in:

- Programming in R and Keras.
- Tackling real-world challenges like overfitting and hyperparameter tuning.
- Applying deep learning techniques to analyze financial data and other domains.

For further information, the course recommends resources such as R's homepage and Keras documentation.