Welcome to the CoGrammar Tutorial: Neural Networks

The session will start shortly...

Questions? Drop them in the chat. We'll have dedicated moderators answering questions.

Data Science Session Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly. **(Fundamental British Values: Mutual Respect and Tolerance)**
- No question is daft or silly - **ask them!**
- There are **Q&A sessions** midway and at the end of the session, should you wish to ask any follow-up questions. Moderators are going to be answering questions as the session progresses as well.
- If you have any questions outside of this lecture, or that are not answered during this lecture, please do submit these for upcoming Academic Sessions. You can submit these questions here: **Questions**

Data Science Session Housekeeping cont.

- For all **non-academic questions**, please submit a query: **www.hyperiondev.com/support**
- Report a **safeguarding** incident: **www.hyperiondev.com/safeguardreporting**
- We would love your **feedback** on lectures: **Feedback on Lectures**

Skills Bootcamp
8-Week Progression Overview

Fulfil 4 Criteria to Graduation

✅ Criterion 1: Initial Requirements

Timeframe: First 2 Weeks
Guided Learning Hours (GLH): Minimum of 15 hours
Task Completion: First four tasks

Due Date: 24 March 2024

✅ Criterion 2: Mid-Course Progress

60 Guided Learning Hours

Data Science - **13 tasks**
Software Engineering - **13 tasks**
Web Development - **13 tasks**

Due Date: 28 April 2024

CoGrammar

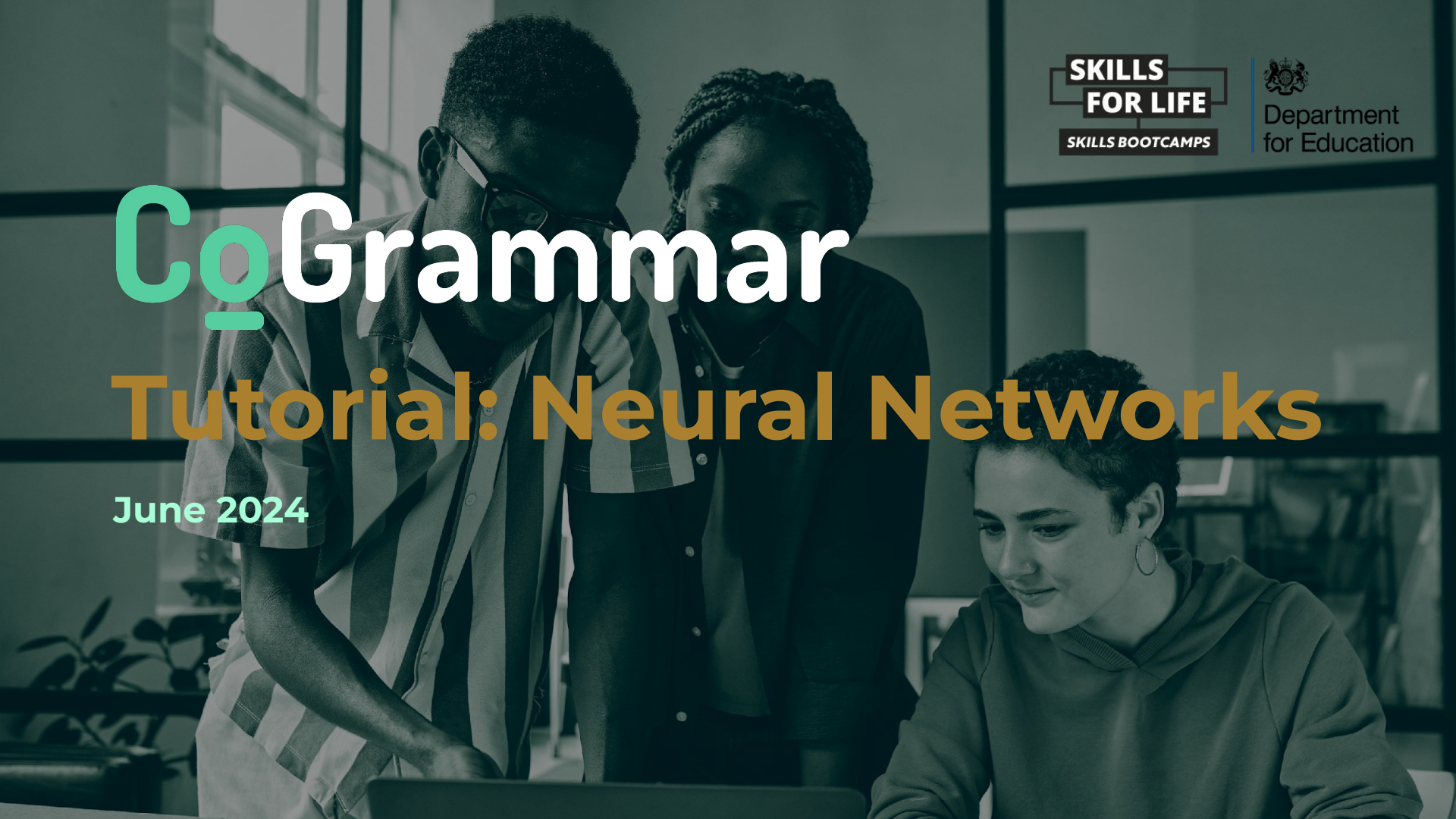
Skills Bootcamp Progression Overview

✅ Criterion 3: Course Progress

Completion: All mandatory tasks, including Build Your Brand and resubmissions by study period end
Interview Invitation: Within 4 weeks post-course
Guided Learning Hours: Minimum of 112 hours by support end date (10.5 hours average, each week)

✅ Criterion 4: Demonstrating Employability

Final Job or Apprenticeship Outcome: Document within 12 weeks post-graduation
Relevance: Progression to employment or related opportunity


SKILLS FOR LIFE
SKILLS BOOTCAMPS
Department for Education
CoGrammar
Tutorial: Neural Networks
June 2024

Important Notice

Please check your spam folders for any important communication from us. If you have accidentally unsubscribed, please reach out to your support team.

Learning Objectives

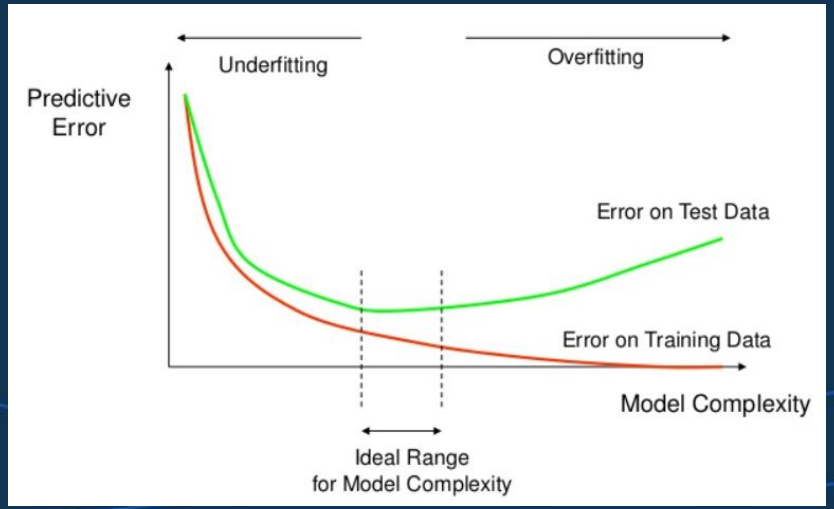
- Understand the fundamental concepts of regularisation
- Know the difference between L1 and L2 regularisation
- Understand the practical considerations for choosing the strength of regularisation
- Understand how the process of hyperparameter tuning greatly influences model quality and performance

Regularisation

CoGrammar

Regularisation

- Overfitting occurs when a model learns the training data too well, creating complex models which capture the noise and fluctuations rather than the underlying pattern.
- Regularisation techniques help to prevent overfitting by adding a penalty to the loss function, discouraging complex models.



CoGrammar

L1 and L2 Regularisation

- The loss function quantifies the discrepancies between actual outcomes (y) and our model's predictions (ŷ) i.e., the error.
- The penalty added to the loss function is proportional to the size of the weights i.e., the larger the weights, the larger the penalty added to the loss function.
- **L1 Regularisation** adds a penalty equal to the sum of the absolute values of the weights.
- **L2 Regularisation** adds a penalty equal to the sum of the squared values of weights.
- **λ** is the regularisation parameter that controls the **strength of the regularisation penalty** applied to the loss function.

$$Loss = Error(y, \hat{y})$$

Loss function with no regularisation

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

Loss function with L1 regularisation

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

Loss function with L2 regularisation

CoGrammar

L1 and L2 Regularisation

Regularization Type	Advantages	Disadvantages
L1	1. Performs feature selection by shrinking less important feature weights to zero.	1. Not effective for datasets with many important features.
	2. Can be used for high-dimensional datasets with many irrelevant features.	2. The solution may not be unique, which can lead to instability in the model.
L2	1. Provides a smooth solution and improves the generalization performance of the model.	1. May not perform well for datasets with many irrelevant features.
	2. Can handle datasets with many important features.	2. Does not perform feature selection, and all features are included in the model with non-zero weights.

Source: https://induraj2020.medium.com/how-does-l1-and-l2-regularization-prevent-overfitting-223ef7001042

Practical Considerations for Choosing λ

- Choosing an appropriate λ (Lambda) is essential for achieving optimal performance.
- λ determines how much to penalise the model for having large weights.
  - Higher values of λ result in stronger regularisation, leading to simpler models with smaller weights
  - If λ is too strong, the model's weights will be too small, leading to underfitting. If λ is too weak, the risk of overfitting will not be adequately reduced.
- Lambda is usually tuned through experimentation or cross-validation to find the optimal balance between bias (models that are too simple) and variance (models that are too complex).

Questions and Answers

CoGrammar

Thank you for attending

SKILLS FOR LIFE
SKILLS BOOTCAMPS

Department for Education

CoGrammar