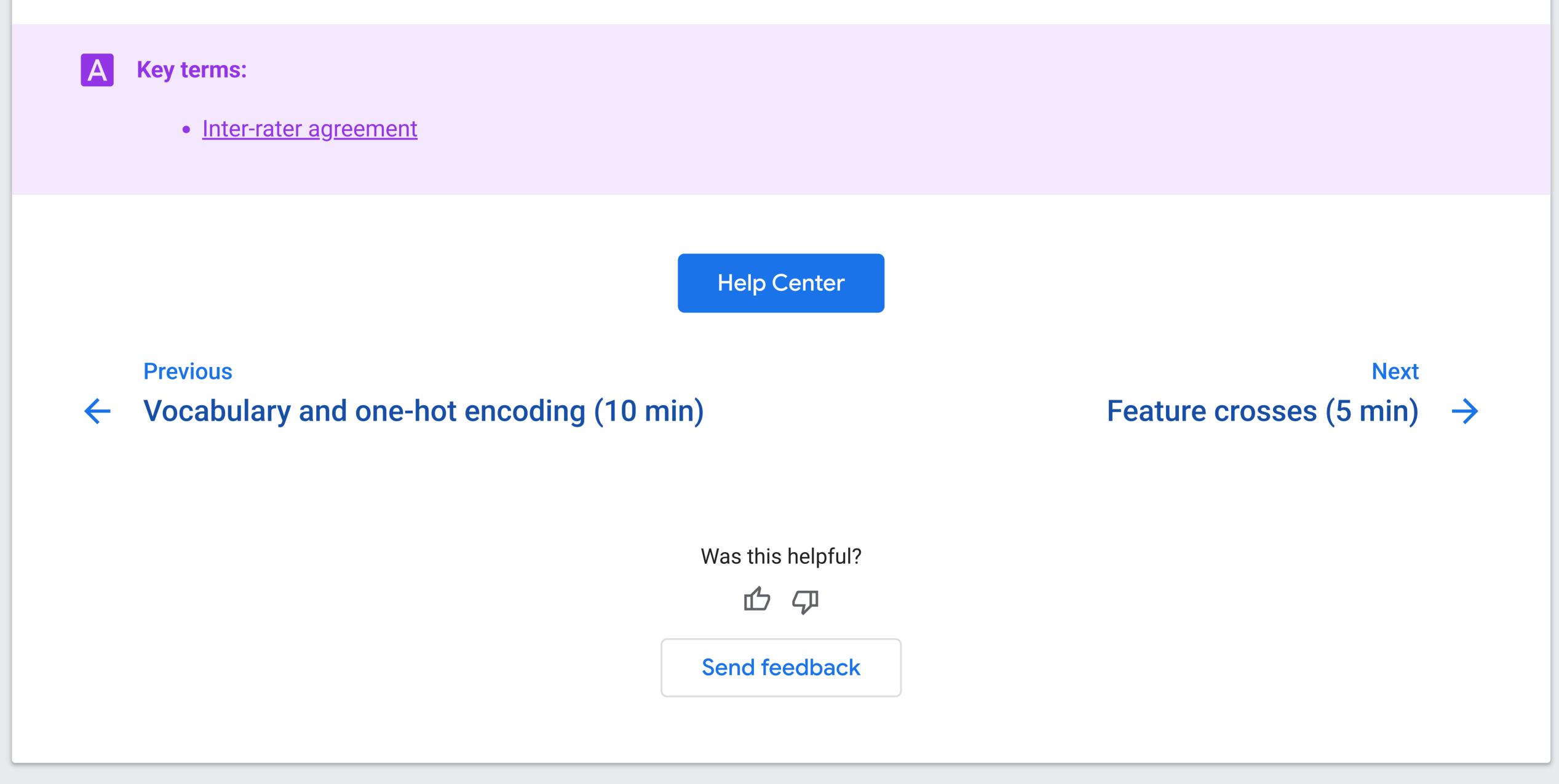


High dimensionality

Categorical data tends to produce high-dimensional feature vectors; that is, feature vectors having a large number of elements. High dimensionality increases training costs and makes training more difficult. For these reasons, ML experts often seek ways to reduce the number of dimensions prior to training.

For natural-language data, the main method of reducing dimensionality is to convert feature vectors to embedding vectors. This is discussed in the Embeddings module later in this course.



Except as otherwise noted, the content of this page is licensed under the Creative Commons Attribution 4.0 License, and code samples are licensed under the Apache 2.0 License. For details, see the Google Developers Site Policies. Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2024-10-18 UTC.

