EE-559 Deep learning – Practice 7, Students' questions

A question is denoted by **Q**; the corresponding answer is denoted by **A**. The questions-related exercises are marked by their numbers in Practice_7.pdf and Practice_7.ipynb documents.

Q: What does the include=objectargument do in dataframe.describe() ?
A: dataframe.describe() "generates descriptive statistics. Descriptive statistics include those that summarize the central tendency, dispersion and shape of a dataset's distribution, excluding NaN values. Analyzes both numeric and object series, as well as DataFrame column sets of mixed data types. The output will vary depending on what is provided." For more details, see the Pandas DataFrame documentation.

To include all columns of the input in the output use include="all". Using include=object will limit the results to object columns. An "object type can hold any Python object, including strings." By default (i.e. include=None) will include all numeric columns in the result. To select pandas categorical columns you can use include="category". However, this only works if the column dtype is set to "category" and not "object".

Categorical data is a type of data in which the variables can take on "a limited, and usually fixed, number of possible values", usually drawn from a predetermined list or category. Some datasets might store the categorical data as dtype objects, since missing values (i.e. NaN) are not included as a category. More about categorical data in pandas is available here: `https://pandas.pydata.org/docs/user_guide/categorical.html`. To check the datatypes for each column in a dataframe at once, you can use dataframe.dtypes.

Detailed discussions about pandas dtypes can be found here:

- `https://stackoverflow.com/questions/29245848/what-are-all-the-dtypesthat-pandas-recognizes`

- `https://stackoverflow.com/questions/29803093/check-which-columns-indataframe-are-categorical`

- `https://stackoverflow.com/questions/29245848/what-are-all-the-dtypesthat-pandas-recognizes`

- `https://stackoverflow.com/questions/48256395/when-should-i-use-pandascategorical-dtype`

Q: Is there a way to reload modules in jupyter notebook without having to restart the runtime?
A: You can use autoreload to reload modules before executing the code. To do so, add the following instructions at the top of the jupyter notebook:

%load_ext autoreload
%autoreload 2