**About mouse dataset:**

The data set consists of a subset of the Mice Protein Expression dataset on Kaggle. The subset contains expression levels of 15 proteins/protein modifications that produce detectable signals in the nuclear fraction of cortex. There are 8 control mice and 8 trisomic mice (Down syndrome), for a total of 16 mice. In the experiments, 2 measurements are present for each mouse and each protein. Therefore, there are 8x2, or 16 measurements, for control mice and trisomic mice. The dataset contains a total of 32 measurements per protein. Each measurement can be considered as an independent sample/mouse.

The eight classes of mice are described based on features such as genotype, behavior and treatment. According to genotype, mice can be ‘control’ or ‘trisomic’. According to behavior, some mice have been stimulated to learn (context-shock) and others have not (shock-context) and in order to assess the effect of the drug memantine in recovering the ability to learn in trisomic mice, some mice have been injected with the drug and others have not.

Classes:

1. c-CS-s: control mice, stimulated to learn, injected with saline (9 mice)
2. c-CS-m: control mice, stimulated to learn, injected with memantine (10 mice)
3. c-SC-s: control mice, not stimulated to learn, injected with saline (9 mice)
4. c-SC-m: control mice, not stimulated to learn, injected with memantine (10 mice)
5. t-CS-s: trisomy mice, stimulated to learn, injected with saline (7 mice)
6. t-CS-m: trisomy mice, stimulated to learn, injected with memantine (9 mice)
7. t-SC-s: trisomy mice, not stimulated to learn, injected with saline (9 mice)
8. t-SC-m: trisomy mice, not stimulated to learn, injected with memantine (9 mice)

### Attribute Information

[1] Mouse ID

[2:16] Values of expression levels of 15 proteins; the names of proteins are followed by N indicating that they were measured in the nuclear fraction. *For example: DYRK1A\_n*

[79] Genotype: control (c) or trisomy (t)

[80] Treatment type: memantine (m) or saline (s)

[81] Behavior: context-shock (CS) or shock-context (SC)

[82] Class: c-CS-s, c-CS-m, c-SC-s, c-SC-m, t-CS-s, t-CS-m, t-SC-s, t-SC-m

**Template question:**

The aim is to identify subsets of proteins that are discriminant between the classes.

**References:**

Original dataset on Kaggle: <https://www.kaggle.com/ruslankl/mice-protein-expression/version/1?select=Data_Cortex_Nuclear.csv>

Source: [UC Irvine Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression)