Real World Examples

1. **Blog Feedback prediction**

   This data originates from blog posts. The raw HTML-documents of the blog posts were crawled and processed. The regression task associated with the data is the prediction of the number of comments in the upcoming 24 hours. In order to simulate this situation, we choose a base-time (in the past) and select the blog posts that were published at most 72 hours before the selected base date/time. Then, we calculate all the features of the selected blog posts from the information that was available at the basetime, therefore each instance corresponds to a blog post. The target is the number of comments that the blog post received in the next 24 hours relative to the basetime.

   Link to Dataset: `http://archive.ics.uci.edu/ml/datasets/BlogFeedback`

2. **Breast Cancer Wisconsin (Prognostic) Data Set**

   Predict the recurrence time/disease-free time of the breast cancer patients from the first 30 features computed from a digitized image of a fine needle aspirate (FNA) of a breast mass.

   Link to Dataset: `http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Prognostic%29`

3. **Multiple Output Regression**

   Until now, we mainly concentrated on a single target variable from the input vector. In this example, we use supply chain management data (scm1d) to predict target vector from the input vector. A brief description of multioutput regression can be found in Section 3.1.5 of Bishop's book.

Link to Dataset: https://osdn.net/projects/sfnet_mulan/downloads/datasets/multi-target%20regression%20datasets/scm1d-train.zip/

Dataset Description: https://arxiv.org/pdf/1211.6581.pdf