Multi-label classification

Multi-label classification is a machine learning task that involves predicting multiple binary target variables (called 'labels') $Y_1 \in \{0,1\}, Y_2 \in \{0,1\}, \ldots, Y_K \in \{0,1\}$ simultaneously based on a feature vector x.

Example: Binary classification/Multi-class classification/Multi-label classification:

- Binary classification: Is there a mountain in the image (**yes**, no)?
- Multi-class classification: which object does the photo show: **a mountain**, a desert, or a car?
- Multi-label classification: Is there a **mountain**, **trees**, **forests**, lakes, sea, **glaciers** in the image? In this case label vector is $Y = (1, 1, 1, 0, 0, 1)$.

Example: Medical diagnosis (a patient can have multiple conditions)

Tasks (Lab 10):

1. Download multi-label dataset emotions:
   `X, Y = fetch_openml("emotions", version=4, return_X_y=True)`

2. Compare the performance of 3 methods:

   (a) BR (Binary Relevance) which involves fitting independent classification models for each binary target variable: $Y_1, \ldots, Y_K$. Each model estimates posterior probability:

   $$P(Y_k = 1 | X = x), \text{ for } k = 1, \ldots, K.$$

   (b) CC (Classifier chains), which involves fitting a chain of models:

   - Model 1: $Y_1 \leftarrow x$
   - Model 2: $Y_2 \leftarrow x, Y_1$
   - Model 3: $Y_3 \leftarrow x, Y_1, Y_2$
   - ...
   - Model K: $Y_K \leftarrow x, Y_1, Y_2, \ldots, Y_{K-1}$

   Analyze the impact of the order of fitting the models in the chain on their effectiveness.

   (c) ECC (Ensemble of Classifier chains), which is based on averaging CC models for different model fitting orders.

3. Use different evaluation scores:

   (a) Subset accuracy: $I(y = \hat{y})$, averaged over all observations in testing data.

   (b) Hamming score: $\frac{1}{K} \sum_{k=1}^{K} I(\hat{y}_k = y_k)$, averaged over all observations in testing data.

   In the above formulas $y = (y_1, \ldots, y_K)$ and $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_K)$.