Data Description for Breast Cancer Data

Each record represents follow-up data for one breast cancer case. These are patient cases exhibiting indications of breast cancer and no evidence of distant metastases at the time of diagnosis.

The prediction problem is to build a reliable and powerful model to predict malignant breast cancers and to interpret the prediction model to identify markers that can be used in practice by physicians to identify patients more likely to have malignant cancers.

During the course, you will be learning different models and your goal will be to use the different models as competing predictions so you can compare how different models work and function. In each assignment, you will be making a recommendation for the prediction model you propose ("reliable and powerful") and interpreting the model for problem solution and practice application.

The number of input variables, or attributes for each sample, is 9. All input variables are numeric-valued and represent measurements from digitized histopathological images of a fine-needle aspiration (FNA) biopsy. The target variable is diagnose as benign (non-cancerous; 0) or malignant (cancerous, 1), respectively. The following list summarizes the variables information:

- **clump_thickness**: (1-10). Benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayers.
- **cell_size_uniformity**: (1-10). Cancer cells tend to vary in size and shape.
- **cell_shape_uniformity**: (1-10). Cancer cells tend to vary in shape and size.
- **marginal_adhesion**: (1-10). Normal cells tend to stick together while cancer cells tend to lose this ability, so the loss of adhesion is a sign of malignancy.
- **single_epithelial_cell_size**: (1-10). It is related to the uniformity mentioned above. Epithelial cells that are significantly enlarged may be a malignant cell.
- **bare_nuclei**: (1-10). This is a term used for nuclei not surrounded by cytoplasm (the rest of the cell). Those are typically seen in benign tumors.
- **bland_chromatin**: (1-10). Describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells, the chromatin tends to be more coarse and to form clumps.
- **normal_nucleoli**: (1-10). Nucleoli are small structures seen in the nucleus. In normal cells, the nucleolus is usually very small if visible at all. The nucleoli become more prominent in cancer cells, and sometimes there are multiple.
- **mitoses**: (1-10). Cancer is essentially a disease of uncontrolled mitosis.
- **diagnose (variable name: NCLASS)**: (0 or 1). Benign (non-cancerous) or malignant (cancerous) lump in a breast.