Gene Expression Classification

Y.A. Qi, T.P. Minka, R.W. Picard, and Z. Ghahramani

**Task** Classify gene expression datasets into different categories, f.g., normal v.s. cancer.

**Challenge** Thousands of genes are measured in the micro-array data, while only a small subset of genes are believed to be relevant for the classification tasks.

**Approach** Predictive Automatic Relevance Determination. This method brings Bayesian tools to bear on the problem of selecting which genes are relevant.

---

**Classifying Leukemia Data**

**Task** Distinguish acute myeloid leukemia (AML) from acute lymphoblastic leukemia (ALL).

**Data** 47 and 25 samples of type ALL and AML respectively with 7129 features per sample. The dataset was randomly split 100 times into 36 training and 36 testing samples, with the new method run on each.

---

**Classifying Colon Cancer**

**Task** Discriminate tumor from normal tissues using microarray data.

**Data** 22 normal and 40 cancer samples with 2000 features per sample. We randomly split the dataset into 50 training and 12 testing samples 100 times and run the new method on each.