Hypothesis Testing and Classification

We denote the set of possible hypothesis $\mathcal{H} = \{H_0, H_1, \dots, H_{M-1}\}$

- Each proposes a different model for the observed data
- This is refered to as the **model family**
- Hypoithesis testing can be equivalently interpreted as classification
- H_0 is sometimes referred to as the null hypothesis

The observed data is a random vector $\mathbf{y}$

In the Bayesian approach, a complete description of our models includes:

- *a priori* probabilities: $p_{\mathcal{H}}(H_m)$
- The conditional probability distributions under each hypothesis: $p_{\mathbf{y}|\mathcal{H}}(\cdot|H_m)$

Our complete characterization of our knowledge of the correct hypothesis is the *a posteriori* probabilities:

$$p_{\mathcal{H}|\mathbf{y}}(H_m|\mathbf{y})$$

This is computed the Bayes' Rule:

$$p_{\mathcal{H}|\mathbf{y}}(H_m|\mathbf{y}) = \frac{p_{\mathbf{y}|\mathcal{H}}(\mathbf{y}|H_m)p_{\mathcal{H}}(H_m)}{\sum_{m'} p_{\mathbf{y}|\mathcal{H}}(\mathbf{y}|H_m)p_{\mathcal{H}}(H_m)}$$

Binary Hypothesis Testing

In the binary case, the model consists of:

- Prior probabilities P_0, P_1
- Likelihood functions $p_{y|H}(y|H_0), p_{y|H}(y|H_1)$

The solution to this hypothesis test is specified as a decision rule, or a classifier

- It maps every possible observation y to one of the two hypotheses
- It partitions the observation space into disjoint decision regions

To quantify what is the best, we use an objective function that defines the loss / cost

- $C(H_j, H_i) = C_{ij}$
- Denotes the cost of deciding the hypothesis is H_i when the correct is H_j
- Our goal is then to find the hypothesis that minimizes the cost $\phi(f)$
  - $\phi(f) = \mathbb{E}\left[C(H, f(y))\right]$
  - Known as the Bayes risk
  - Expectation is takenover random y and H
- The costs C_{ij} are **valid** if the cost of a correct decision is lower than that of an incorrect decision

Likelihood Ratio Test

For a binary test between H_0 and H_1, we calculate:

$$L(y) = \frac{p_{y|H}(y|H_1)}{p_{y|H}(y|H_0)}$$

$$\eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

Choose H_1 when $L(y) > \eta$ and H_0 when $L(y) < \eta$ and arbitrarily otherwise

- $L(y)$ is referred to as the likelihood ratio

  - Note that this is computed exclusively from the model and the data
  - It is a statistic, a real-valued function of the data
  - This theorem tells us that $L(y)$ summarizes the data to tell us everything we need to make the best decision about H
    - That is, it is a sufficient statistic
    - Any invertible function of the likelihood function is a sufficient statistic
  - This is a random variable of the data
- η is computed exclusively from the apriori probabilities and costs
  - It can be thought of as a bias based on the costs and apriori probabilities for how extreme our data needs to be

Proof Sketch:

- The main idea is to realize that $C_{00}p_{H|y}(H_0|y) + C_{01}p_{H|y}(H_1|y)$ represents the conditional expectation of the Bayes risk if we choose H_0
- We want to choose whichever one leads to lower cost

TODO: 2.2.3 onwards